Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGE-VAR, and VAR Models

Anders Warne, Günter Coenen and Kai Christoffel*

December 8, 2015

Summary: The predictive likelihood is useful for ranking models in forecast comparison exercises using Bayesian inference. We discuss how it can be estimated, by means of marginalization, for any subset of the observables in linear Gaussian state-space models. We compare macroeconomic density forecasts for the euro area of a DSGE model to those of a DSGE-VAR, a BVAR, and a multivariate random walk over 1999Q1–2011Q4. While the BVAR generally provides superior forecasts, its performance deteriorates substantially with the onset of the Great Recession. This is particularly notable for longer-horizon real GDP forecasts, where the DSGE and DSGE-VAR models perform better.

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

1. Introduction

In Bayesian analysis of time series, the predictive likelihood can be employed to compare forecast accuracy across models. 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."

Note: This paper has previously been circulated with the titles "Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models" and "Forecasting with DSGE-VAR Models". We are particularly grateful for comments from the editor and two referees, and to Marta Bańbura who specified the BVAR model we use in the paper. We are also very grateful for discussions with Gianni Amisano (ECB and Board of Governors), Michal Andrle (IMF), Jan Brůha (Czech National Bank), Herman van Dijk (Tinbergen Institute), Juha Kilponen (Suomen Pankki), Bartosz Maćkowiak (ECB), Frank Schorfheide (University of Pennsylvania), Mattias Villani (Linköping University), and comments from members of the Working Group on Econometric Modelling, and participants of the Tinbergen Institute workshop on "Recent Theory and Applications of DSGE Models" at Erasmus University Rotterdam in June 2012, 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, Sonnemannstrasse 20, 60314 Frankfurt am Main, Germany.
Corresponding author: Anders Warne, e-mail: anders.warne@ecb.europa.eu, phone: +49-6913448737.

A problem which often occurs in such forecast comparison exercises is that the investigator is interested in comparing the performance with respect to some but not all of the observable variables that can be predicted. Before a meaningful comparison can be made, the observables which are not regarded as interesting must therefore be integrated out from the predictive likelihood of the models where they appear.

This marginalization problem may be solved by applying textbook results when the joint predictive likelihood of a model has a known distributional form. However, models with a known predictive distribution are rare and in the typical cases when this distribution is unknown, we can instead make use of the fact that the predictive likelihood is equal to the integral of the conditional likelihood (the predictive likelihood *conditional* on a value for the parameters) times the posterior density with respect to the model parameters. If the conditional likelihood is based on a distribution where marginalization can be handled analytically, then the marginalization problem can be solved at this stage.

In this paper we suggest a recursive approach, based on the Kalman filter, to marginalize the conditional likelihood in linear Gaussian discrete-time state-space models. Our approach builds up the marginalized parts of only the relevant arrays and is therefore—except when dealing with one-step-ahead forecasts—simpler than first calculating the mean and the covariance matrix of the joint conditional distribution and thereafter reducing these arrays to the entries relevant for the marginalized conditional likelihood; see e.g. Andersson and Karlsson (2008) and Karlsson (2013, Section 8.2.1).[1]

We apply the recursive approach to marginalizing the predictive likelihood to the forecast comparison exercise in Christoffel, Coenen and Warne (2011)—henceforth, CCW. In this study, we followed the normal approximation approach suggested by Adolfson, Lindé and Villani (2007) for computing the marginalized predictive likelihood[2] when comparing

[1] The Kalman filter based approach suggested in our paper has been applied in a more recent study by Del Negro, Hasegawa and Schorfheide (2014).

[2] In their empirical forecasting study, Adolfson, Lindé and Villani (2007) discuss this marginalization problem in the context of marginal likelihoods. Specifically, the authors first note that the joint predictive

the density forecasts of the ECB's New Area-Wide Model (NAWM) to various alternatives.[3] This pseudo out-of-sample forecast comparison exercise covered the period after the introduction of the euro and ended in 2006Q4, focusing on three nested partitions of the 12 (out of 18) observable variables that are endogenously determined in the NAWM.

In the novel application of the current paper, we enhance the forecast comparison exercise in CCW in several ways. First, while the forecast sample begins in 1999Q1, as in CCW, the end point is extended from 2006Q4 to 2011Q4. This allows us to study how the models compare in terms of forecasting performance during and after the onset of the Great Recession. Second, we assess the results from using the normal approximation of the predictive likelihood to those obtained from an estimator of the predictive likelihood based on Monte Carlo integration of the marginalized conditional likelihood with respect to posterior draws of the model parameters. Furthermore, we include in this setting a DSGE-VAR with the NAWM as prior and compare the density forecast performance of the DSGE model to this DSGE-VAR model, as well as to a BVAR from CCW, and a multivariate random walk model estimated with Bayesian methods. It should be stressed that all these models are estimated on the same set of variables.

The remainder of the paper is organized as follows. Section 2 introduces notation and presents concepts related to the predictive likelihood. Moreover, it considers linear state-space models with Gaussian innovations and shows how the conditional likelihood can

likelihood (for all observables in a model over $T+1$ until $T+h$) is equal to the ratio of the marginal likelihood of the historical sample (up to time T) and the forecast sample (between time $T+1$ and $T+h$), and the marginal likelihood of the historical sample only; see equation (5.3) in Adolfson, Lindé and Villani. While the marginal likelihood for the whole sample may be decomposed into a sequence of intermediate non-overlapping joint predictive likelihoods, they also mention that the marginal likelihood cannot be decomposed into terms of the marginal h-step-ahead predictive likelihood when $h > 1$. They therefore conclude that the marginal likelihood cannot detect whether some models perform well on certain forecast horizons while other models do better on other horizons; see the last paragraph on page 324 of Adolfson, Lindé and Villani. Moreover, since the number of variables included in their density forecast comparison exercise is large, the authors claim that kernel density estimation of the predictive likelihood is not practical, and instead they assume that the predictive likelihood for marginal h-step-ahead forecasts is multivariate normal and estimate the mean vector and the covariance matrix for the predictive sample.

[3] See Christoffel, Coenen and Warne (2008) for details about the NAWM, an open-economy DSGE model of the euro area.

be marginalized via a Kalman filter. Section 3 presents the empirical density forecast comparison exercise, while Section 4 summarizes the main findings of the paper.

2. The Predictive Likelihood

To establish notation, let $\theta_m \in \Theta_m$ be a vector of unknown parameters of a model indexed by m, while $\mathcal{Y}_T = \{y_1, y_2, \ldots, y_T\}$ is a real-valued time series of an n-dimensional vector of observables y_t. The observed value of this vector of random variables is denoted by y_t^o, while the sample of observations is similarly denoted by $\mathcal{Y}_T^o$. The observables' density function 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, $p(\theta_m|\mathcal{Y}_T^o, m)$.

Point and density forecasts are determined from the predictive density of model m. For a sequence of future values of the observable variables $\mathcal{Y}_{T,h} = \{y_{T+1}, \ldots, y_{T+h}\}$, this density can be expressed as

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

where T increases as additional data points are added to the historical sample in a recursive forecast exercise. The joint predictive likelihood of model m is equal to the predictive density in (1) evaluated at the observed values $\mathcal{Y}_{T,h}^o = \{y_{T+1}^o, \ldots, y_{T+h}^o\}$.

Suppose we are interested in forecasting a subset of the variables $\mathcal{Y}_{T,h}$, denoted by $\mathcal{Y}_{s,T,h} = \{y_{s,T+1}, \ldots, y_{s,T+h}\}$, where $y_{s,T+i} = S_i' y_{T+i}$, and S_i is an $n \times n_i$ selection matrix with $n_i \in \{0, 1, \ldots, n\}$ for $i = 1, \ldots, h$. The subset of variables in $y_{s,T+i}$ is the same for each time period $T + i$ when i is fixed and T varies, while the subset of variables can vary when T is fixed and i varies.[4]

[4] Hence, S_i does not need to be the same as *or* have the same dimension as S_j when $i \neq j$, with the consequence that different subsets of observables may be included in $y_{s,T+i}$ and $y_{s,T+j}$.

The marginalized predictive density of $\mathcal{Y}_{s,T,h}$ can be expressed as

$$p\big(\mathcal{Y}_{s,T,h}\big|\mathcal{Y}_T^o, m\big) = \int_{\Theta_m} p\big(\mathcal{Y}_{s,T,h}\big|\mathcal{Y}_T^o, \theta_m, m\big)p\big(\theta_m\big|\mathcal{Y}_T^o, m\big)d\theta_m. \tag{2}$$

The marginalized predictive likelihood is given by (2) when evaluated at the observed values $\mathcal{Y}_{s,T,h}^o$, while the term $p\big(\mathcal{Y}_{s,T,h}^o\big|\mathcal{Y}_T^o, \theta_m, m\big)$ is called the marginalized *conditional* likelihood.

Below we suggest an approach to marginalizing the conditional likelihood which applies to linear Gaussian state-space models and which is simple, fast, and robust. Once this likelihood has been determined, the problem of calculating the marginalized predictive likelihood in (2) for $\mathcal{Y}_{s,T,h}^o$ depends on selecting an appropriate numerical method for integrating out the dependence on the parameters.

2.1. Marginalization in Linear Gaussian State-Space Models

Standard linear time series models—including VAR models, VARMA models, dynamic factor models, and other unobserved component models—may be cast in state-space form. Structural models, such as log-linearized DSGE models and other linear rational expectations models, also have such a representation provided that a unique and convergent solution exists for a given value of the model parameters.

To establish some further notation, let the observable variables y_t be linked to a vector of state variables ξ_t of dimension r through the equation

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

The errors, w_t, are assumed to be i.i.d. $N(0, R)$, with R being an $n \times n$ positive semidefinite matrix, while the state variables are determined from a first-order VAR system:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1, \ldots, T. \tag{4}$$

The state shocks, η_t, are of dimension q and i.i.d. $N(0, I_q)$ and independent of w_τ for all t and τ, while F is an $r \times r$ matrix, and B is $r \times q$. The parameters of this model, (μ, H, R, F, B), are uniquely determined by θ_m.

The system in (3) and (4) is a state-space model, where equation (3) gives the measurement or observation equation and (4) the state or transition equation. Provided that the number of measurement errors and state shocks is large enough and an assumption about the initial conditions is added, we can calculate the likelihood function with a suitable Kalman filter; see, e.g., Harvey (1989) or Durbin and Koopman (2012).

Suppose we are interested in forecasting the subset of variables $\mathcal{Y}_{s,T,h}$ with the state-space system and that we have the observed values $\mathcal{Y}^o_{s,T,h}$. Let $\mu_{s,i} = S_i'\mu$, $H_{s,i} = HS_i$, and $R_{s,i} = S_i'RS_i$ when $n_i \geq 1$ for $i = 1, \ldots, h$. We here find for $h \geq 1$

$$\log p\big(\mathcal{Y}^o_{s,T,h} \big| \mathcal{Y}^o_T, \theta_m, m\big) = \sum_{i=1}^{h} \log p\big(y^o_{s,T+i} \big| \mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big), \tag{5}$$

where $\mathcal{Y}^o_{s,T,0}$ is empty by definition. If $n_i \geq 1$

$$\begin{aligned}\log p\big(y^o_{s,T+i} \big| \mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big) &= -\frac{n_i}{2}\log(2\pi) - \frac{1}{2}\log\left|\Sigma_{y_s,T+i|T+i-1}\right| \\ &-\frac{1}{2}\left(y^o_{s,T+i} - y_{s,T+i|T+i-1}\right)' \Sigma^{-1}_{y_s,T+i|T+i-1}\left(y^o_{s,T+i} - y_{s,T+i|T+i-1}\right),\end{aligned} \tag{6}$$

where

$$y_{s,T+i|T+i-1} = \mu_{s,i} + H'_{s,i}\xi_{T+i|T+i-1},$$

$$\Sigma_{y_s,T+i|T+i-1} = H'_{s,i}P_{T+i|T+i-1}H_{s,i} + R_{s,i}.$$

The vector $\xi_{T+i|T+i-1}$ is the one-step-ahead forecast of the state variables, while $P_{T+i|T+i-1}$ is the corresponding forecast error covariance matrix. The log of the marginalized conditional log-likelihood value at $T + i$ is zero for $n_i = 0$; see also the online appendix (Appendix C) for further details.

The above approach provides a *bottom-up* evaluation of the marginalized conditional likelihood. Since the joint conditional likelihood for $\mathcal{Y}^o_{T,h}$ is normal, the marginalized conditional likelihood for $\mathcal{Y}^o_{s,T,h}$ is also normal with mean and covariance obtained by selecting the appropriate elements from the mean vector and covariance matrix of the joint predictive distribution conditional on the parameters. This property is used by, for instance, Andersson and Karlsson (2008), when estimating the marginalized predictive likelihood for VAR models via the conditional likelihood. When h is large, such a *top-down* approach to evaluating the marginalized conditional likelihood is expected to be slower than the bottom-up Kalman filter evaluation presented above, especially when the evaluations are repeated many times. Note also that the bottom-up and top-down approaches coincide only in the case of one-step-ahead forecasts.

2.2. Estimating the Marginalized Predictive Likelihood

Once the problem of evaluating the marginalized conditional likelihood has been addressed, we proceed with the second step in the estimation of the marginalized predictive likelihood, which involves integrating out the dependence on the parameters. It is assumed below that an ergodic sequence of parameter draws are available from the posterior distribution.

In the empirical application in Section 3 we apply a simple Monte Carlo (MC) integration method to estimate the predictive likelihood. Specifically, we use:

$$\hat{p}_{MC}\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, m\big) = \frac{1}{N}\sum_{j=1}^{N} p\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta^{(j)}_m, m\big), \tag{7}$$

where $\theta^{(j)}_m$ is a draw from the posterior density $p(\theta_m|\mathcal{Y}^o_T, m)$ for $j = 1, \ldots, N$. Under certain regularity conditions (Tierney, 1994), the right hand side of (7) converges almost surely to the expected value of $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta_m, m)$ with respect to $p(\theta_m|\mathcal{Y}^o_T, m)$, i.e. to the predictive likelihood $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, m)$. Hence, equipped with the posterior draws, the marginalized predictive likelihood can be consistently estimated from the sample average of the marginalized conditional likelihood. A further property of this estimator is that it is unbiased (see Chan and Eisenstat, 2015, Proposition 1).

The MC estimator in (7) is expected to work well in practise when the posterior draws cover well enough the parameter region where the marginalized conditional likelihood is large. This is more likely to be the case when the dimension of $\mathcal{Y}_{s,T,h}$ is fairly small *and* h is not too large, but it is not obvious when one or both of these properties is not met. Clearly, when $S_i = I_n$ for all h and the latter is sufficiently large, the situation resembles the case where the marginal likelihood is estimated by averaging the likelihood over draws from the prior distribution, and such an estimator is expected to be poor. Standard methods for calculating the marginal likelihood (e.g., Geweke, 2005, and Sims, Waggoner and Zha, 2008) may instead be considered, but this is beyond the scope of the current paper.

3. Comparing Forecast Accuracy: An Application to Euro Area Data

In this section we compare marginalized h-step-ahead density forecasts for three subsets of observables across four linear Gaussian state-space models for euro area data using the approach discussed above. In Section 3.1 we discuss the models: a DSGE model, a DSGE-VAR model, a BVAR model, and a multivariate random walk model, where all models have the same observable variables. In Section 3.2, we present the forecast sample and summarize the empirical results of the exercise. The calculations below have to a large extent been performed with the help of YADA, a Matlab program for Bayesian estimation and evaluation of DSGE and DSGE-VAR models; the code and the documentation for YADA may be downloaded from `http://www.texlips.net/yada/`.

3.1. Empirical Models

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

Over recent years, significant efforts have been undertaken to bring DSGE models to the forecasting arena with promising results; see, for example, CCW, Del Negro and

Schorfheide (2013), Smets, Warne and Wouters (2014), Wolters (2015), and references therein. In our application, we extend the forecasting comparison exercise in CCW using the New Area-Wide Model (NAWM) of the ECB. With E_t being the rational expectations operator, the NAWM, like other log-linearized DSGE models, can be written as:

$$A_{-1}\xi_{t-1} + A_0\xi_t + A_1 E_t\xi_{t+1} = D\eta_t, \quad t = 1, 2, \ldots, T. \tag{8}$$

The matrices A_i $(r \times r)$, with $i = -1, 0, 1$, and D $(r \times q)$ are functions of the vector of DSGE model parameters. Provided that a unique and convergent solution of the system (8) exists, we can express the model as the first order VAR system in (4).

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.

Christoffel et al. (2008) adopt the empirical approach outlined in Smets and Wouters (2003) and An and Schorfheide (2007) and estimate the log-linear NAWM with Bayesian methods, using time series for 18 macroeconomic variables. The estimation sample in Christoffel et al. (2008) is given by the period 1985Q1 until 2006Q4, with 1980Q2–1984Q4 serving as training sample.

In our application, we extend the sample to 2011Q4, also taking into account the changes in country composition of the euro area, and reestimate the model with the new data, using the same training sample as in CCW. The time series for the sample 1985Q1–2011Q4 are displayed in Figure E.1 in the online appendix; see Christoffel et al. (2008) and CCW for details on the observed variables and their transformations.

3.1.2. A DSGE-VAR Model with the NAWM as Prior

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 observable variables be treated as endogenous. The VAR model of the NAWM observables y_t can be written as:

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

where $\epsilon_t \sim N(0, \Sigma_\epsilon)$ and with Σ_ϵ being an $n \times n$ positive definite matrix. 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 of y_t exists for $t = 0, \ldots, 1 - p$.

The 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 unknown parameters. BVAR models typically include a number of hyperparameters that are calibrated by the researcher and are therefore included in the model index m rather than among θ_m.[5]

A DSGE-VAR, suggested by Del Negro and Schorfheide (2004), is a well-known example of a VAR model which includes additional unknown parameters through the prior. Specifically, they assume that the prior distribution of the VAR model parameters is conditional on the DSGE model parameters such that the conditional moments of y_t are determined through the implied first and second population moments for a given value of the DSGE model parameters. A prior distribution for *all* parameters is thereafter obtained by multiplying this conditional prior by the marginal prior for the DSGE model parameters.

In addition, DSGE-VARs are indexed by the parameter λ, which determines the weight on the prior relative to the data. The VAR approximation of the DSGE model resides

[5] See Giannone, Lenza and Primiceri (2015) for a novel approach to making inference about the informativeness of the prior distribution of BVARs.

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. Consequently, the DSGE-VAR provides an approach for assessing the degree of misspecification, by relaxing the cross-equation restrictions of the DSGE model, and where lower values of λ suggest a greater degree of misspecification; see Del Negro, Schorfheide, Smets and Wouters (2007).

In our application, the DSGE-VAR model is taken from Warne, Coenen and Christoffel (2013) and has the largest marginal likelihood among all pairs (λ, p) that they considered, i.e. we let $(\lambda, p) = (2.5, 2)$.[6,7]

3.1.3. BVAR and Random Walk Models

We also consider a Bayesian VAR model for the same observable variables y_t as in 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. Below, we employ the BVAR model from CCW, estimated using the methodology in Bańbura, Giannone and Reichlin (2010). This approach relies on using dummy observations when implementing the normal-inverted

---

[6] The common unit-root technology trend in the NAWM implies a number of cointegration relations among the levels of the observables, such as the consumption-output and investment-output ratios, or the wage share. In principle, it may be worthwhile to include these theoretically founded cointegration relations in the VAR model, as in Del Negro et al. (2007), thus forming a vector error correction model (VECM) instead of a VAR. However, for the euro area the cointegration relations implied by the NAWM are typically trending. This suggests that adding the cointegration relations will result in a worse fit for a DSGE-VECM than for a DSGE-VAR; see Canova (2014) for a treatment on the issue of variable transformations in the context of DSGE models. Furthermore, treating the DSGE model-based cointegration space as fixed when λ varies, while all other DSGE model parameters enter the DSGE-VECM via the prior is arbitary. With these objections in mind, we have opted to not include a DSGE-VECM in the forecast comparison exercise below. The interested reader may also consider the forecasting exercise in Adolfson, Lindé and Villani (2007) based on euro area data. In their study, the DSGE model which includes the cointegration relations as observables and a Bayesian VECM based on the same cointegration relations lead to inferior forecasting performance compared with models that exclude these long-run relations, although the deterioration is more pronounced for the DSGE model; see also Adolfson, Laséen, Lindé and Villani (2008).

[7] From Tables E.7–E.19 in the online appendix we find that (i) $p = 2$ gives the largest marginal likelihood value over the recursive estimation sample ending in 1998Q4, 1999Q4, ..., 2010Q4, when both λ and p vary; (ii) the posterior mode of λ varies between 4.5 and 2 for $p = 2$ and decreases as the estimation sample length increases; and (iii) the posterior surface for λ is flat around the mode. Hence, our choice of (λ, p) seems reasonable also from a recursive estimation perspective. The interested reader may also consult Warne et al. (2013), where the DSGE-VAR with $(\lambda, p) = (2.5, 2)$ is compared with a DSGE-VAR with $(\lambda, p) = (6, 4)$. Overall, these two models have very similar forecasting performance over the sample 1999Q1–2006Q4.

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, or 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 BVAR is referred to as the model with a mixed prior. A more detailed description of this BVAR is also found in the online appendix (Appendix A).

The final model we shall consider is a multivariate random walk for the vector y_t with the NAWM observables. We employ 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. $N(0, \Omega)$, where Ω is an $n \times n$ positive definite matrix of unknown parameters, and $p(\Omega) \propto |\Omega|^{-(n+1)/2}$. One advantage of this model is that it allows for an analytical determination of the predictive density. For instance, the marginal h-step-ahead predictive density of $y_{s,T+h}$ is given by a n_h-dimensional Student t-distribution with mean $S_h' y_T^o$, covariance matrix

$$\frac{h}{T-n-1} \sum_{t=1}^{T} S_h' \big(y_t^o - y_{t-1}^o\big) \big(y_t^o - y_{t-1}^o\big)' S_h,$$

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

3.2. Density Forecasts

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

$$\mathcal{S}_{T+N_h+h-1}(m) = \sum_{t=T}^{T+N_h-1} \log p(\mathcal{Y}_{s,t,h}^o | \mathcal{Y}_t^o, m), \quad h = 1, \ldots, h^*, \tag{10}$$

where N_h is the number of time periods the h-step-ahead predictive likelihood is evaluated, and where the time subscript of $\mathcal{S}$ is given by the last time period in the evaluation sample. The log predictive score is optimal in the sense that it uniquely determines the model ranking among all local and proper scoring rules; see Gneiting and Raftery (2007) for a survey on 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.

When comparing the density forecasts of the NAWM, the DSGE-VAR, the BVAR, and the multivariate random walk model below we will evaluate the log predictive score in (10) with realizations for different subsets of the observables $\mathcal{Y}^o_{s,T,h} = y^o_{s,T+h}$. Hence, the predictive likelihood for each model and time period is marginalized with respect to the forecast horizon and the variables included in a subset.

Furthermore, we will also consider the recursive average log predictive score, given by

$$\bar{\mathcal{S}}_{T+N_h+h-1}(m) = \frac{\mathcal{S}_{T+N_h+h-1}(m)}{N_h}, \quad N_h = 1, \ldots, T_h, \tag{11}$$

where T_h is the maximum number of time periods which the h-step-ahead predictive likelihood can be evaluated over the forecast sample. This variant of the log score is convenient when displaying the predictive scores of the models recursively.

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 2011Q4. The maximum forecast horizon is eight quarters, yielding 52 quarters with one-step-ahead forecasts and 45 quarters with eight-step-ahead forecasts. We shall only consider forecasts of quarterly growth rates for the variables in first differences, while CCW also studied forecasts of annual growth rates for such variables.

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 consumption. For the remaining 12 variables we examine three nested subsets. The smallest subset is called the *small selection* and given by real GDP growth, GDP deflator inflation, and the short-term nominal interest rate. This selection may be regarded as the minimum set of variables relevant to a meaningful analysis of 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.

3.2.1. Empirical Evidence using the MC Estimator

The log predictive scores based on the MC estimator of the marginal h-step-ahead predictive likelihood are shown in the upper part of Figure 1 for our entire forecast sample from 1999Q1 to 2011Q4, the three variable selections defined above, eight forecast horizons $h = 1, \ldots, 8$, and our four models.[8,9] For the NAWM and the DSGE-VAR model 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 lower correlation between the draws and a sufficiently high estimation accuracy. This procedure yields estimates of the log predictive likelihood that typically are accurate up to and including the first decimal. We discuss the numerical standard errors in the online appendix (Appendix D).[10]

[8] The log predictive scores estimated with the normal approximation are shown in the lower part of Figure 1 and are discussed below in Section 3.2.2.

[9] While the BVAR and the random walk are reestimated each quarter, the NAWM and the DSGE-VAR are both reestimated on an annual basis over the forecast sample. Specifically, the latter two models are reestimated once the end point of the historical sample is Q4 in a given calendar year.

[10] In summary, we find that the numerical precision of the MC estimator of the log predictive likelihood is satisfactory. For very small values of the log predictive likelihood, the precision falls, especially for models

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 MC estimator. In the case of the random walk model, the predictive likelihood for a selection of variables is multivariate Student t and can therefore be computed from its analytical expression.

When comparing the NAWM with the DSGE-VAR, it is noteworthy that the latter model generally obtains higher log scores for all horizons and variable selections. At the longer horizons, the NAWM obtains values that are near those of the DSGE-VAR and, in the case of the small selection, even slightly higher for the eight-quarter-ahead forecasts. Hence, it seems that taking into account possible misspecification of the NAWM through a DSGE-VAR improves forecasting performance, especially at the shorter horizons.

Compared with the BVAR model, the NAWM is outperformed for the large and medium selections and all forecast horizons, while in the case of the small selection the forecast performance of the BVAR deteriorates relative to the NAWM (and the DSGE-VAR) as the forecast horizon increases. In fact, in this case the BVAR performs worse than the NAWM for the five- to eight-quarter-ahead density forecasts. At the same time, the BVAR is the best performer over all horizons when using the large and the medium selection, with the DSGE-VAR in second place. However, it performs worse than the DSGE-VAR for the small selection and all forecast horizons beyond two quarters.

It is also worth pointing out that the random walk model is competitive with the NAWM for the one-step-ahead forecasts, especially for the small selection. As the forecast horizon increases, however, the random walk model's performance worsens considerably.

CCW identify two main factors that may explain the relative strengths and weaknesses of the NAWM. On the one hand, its explicit microfoundations give rise to a parsimoniously

with a large number of parameters, but it is unlikely that it will impair the validity of the ranking of models since such small values are rare.

parameterized structure with a large number of cross-equation restrictions. This is potentially an advantage for achieving forecast accuracy. On the other hand, the embedded balanced growth path assumption means that the model's ability to deal with differing trends in the observables is limited compared to VAR models. In fact, this assumption may induce a bias in the forecasts and CCW report that this bias is particularly important in the case of variables connected with the wage share. Specifically, the NAWM systematically overpredicts nominal wage growth and underpredicts the private consumption and GDP deflators when the forecast sample ends in 2006Q4. Since real wage growth is systematically overpredicted, real private consumption growth is also overpredicted. These systematic forecast errors primarily affect the NAWM's performance for the large and medium selections, and these findings are also valid in our application where the forecast sample ends in 2011Q4. By relaxing the cross-equation restrictions of the NAWM, the DSGE-VAR model is consequently able to achieve better density forecasts than the NAWM for both of these selections.[11]

In Figure 2 the recursive average log predictive scores until 2011Q4 are displayed for all models based on the large, medium and small selections and the one, two, four, and the eight-quarter-ahead forecasts. For the large and medium selections we find that the ranking of models over the various horizons is not greatly affected by the choice of the sample end point. It is interesting to note that the average log predictive scores of the NAWM and the DSGE-VAR are fairly constant. In view of the Great Recession in late 2008 and early 2009, the drop in forecast performance of these two models is quite small for all the selections.

By constrast, the forecasting performance of the BVAR deteriorates substantially with the onset of the Great Recession. This loss in performance is quite remarkable, especially

[11] Plots of the mean forecast paths from all models and all variables in the large selection are available in Figures E.4–E.9 of the online appendix. Furthermore, the RMSE's based on these forecast paths and using the whole forecast sample are shown in Tables E.4–E.6.

in the case of the small selection with the result that the model looses its first rank position over the longer forecast horizons. In addition, there are several time periods during 2002–2003 where the performance of the BVAR drops for the longer horizons and for all selections. At the onset of both these periods, the largest modulus of the BVAR at the posterior mean increases from values around 0.98 to values above unity. In the first instance, this sign of explosiveness also lasts until 2005, thus covering a period over which the forecasting performance of the BVAR is actually improving.[12]

To gain more insight into possible explanations for the reversal in model ranking for the small selection, we next turn to Figure 3. The focus of the graphs in this figure is a simple decomposition of the log predictive score (likelihood) into a conditional score for GDP deflator inflation and the short-term nominal interest rate given real GDP growth, and the marginal score for real GDP growth. It can be seen from these graphs that the drop in forecasting performance for the BVAR in connection with the Great Recession is mainly explained by a loss in real GDP growth forecasting performance.

To confirm this claim, we also display all the average marginal log predictive scores for the three variables in the small selection in Figure 3. The graphs in this figure show that the loss in forecasting performance of the BVAR at the onset of the Great Recession in 2008Q4 is most severe in the case of real GDP growth.[13]

For the other three models, the average forecast performance of the small selection is also affected by the Great Recession, but here the impact is less striking than for the BVAR. In the case of the NAWM and the DSGE-VAR, it is interesting to note that the performance loss also seems mainly related to the real GDP growth forecasts. In view of the sharp fall in euro area real GDP growth in 2008Q4 and 2009Q1 compared with the behavior of

[12] A graph of the recursive posterior mean based largest modulus values of the BVAR is available in Figure E.2 of the online appendix.

[13] Concerning the loss in performance for the BVAR around 2002-2003, this appears to be mainly related to poor forecasts of GDP deflator inflation, although also for this variable density forecasts of the BVAR drop at the onset of the Great Recession, albeit not by as much as in the case of real GDP growth.

inflation and short-term nominal interest rates, this finding is hardly surprising, but the finding that the "more structural models" are less sensitive than the reduced form BVAR model to such an event is indeed interesting.[14]

3.2.2. Evidence Based on the Normal Approximation

It was suggested by Adolfson, Lindé and Villani (2007) to approximate the marginalized predictive likelihood with a normal density with mean and covariance matrix taken from the predictive density. While such an approximation is not necessary when we know how to estimate the marginalized predictive likelihood, it can nevertheless serve as a tool for enhancing our understanding of the results in Section 3.2.1. In view of the assumption that the conditional likelihood is normal, the normal approximation is a natural reference point as any deviation from normality is due to the impact of the posterior parameter distribution on the predictive likelihood. Consequently, the size of the errors from using a normal approximation relative to the MC estimator is a relevant matter.

The mean and covariance matrix of the predictive density in (1) 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 exist. Let these moments be denoted by $E[\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, \theta_m, m]$ and $C[\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, \theta_m, m]$, respectively. The mean of the predictive density is then given by

$$E\big[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o, m\big] = E_T\Big[E\big[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o, \theta_m, m\big]\Big], \tag{12}$$

[14] This outcome may be explained by the fact that the BVAR prior of the constant term (Φ_0) is diffuse; see the online appendix (Appendix A). By contrast, the constant vector of the NAWM (μ) is calibrated, while the DSGE-VAR has a proper prior for the constant term. In the case of the BVAR, the posterior mean of the constant term for real GDP growth fluctuates around 0.3 before the Great Recession, falls sharply to a minimum close to zero in 2009Q2, and gradually recovers thereafter to about 0.25 at the end of the sample; see Figure E.3 in the online appendix. It is possible that a proper prior for the constant term may render its posterior mean to be less sensitive to the recorded output loss in late 2008 and early 2009 and could therefore result in better density forecasts of real GDP growth over this period. While it may be tempting to change the prior of the BVAR and thereby possibly reduce the sensitivity of its estimated constant term to the Great Recession, it is also conceivable that the diffuse prior has its advantages during the more tranquil periods of the forecast sample. In any event, we prefer to restrict the specification of the BVAR model in our application to follow the methodology of Bańbura, Giannone and Reichlin (2010) for large BVARs.

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

$$C\big[\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, m\big] = E_T\Big[C\big[\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, \theta_m, m\big]\Big] + C_T\Big[E\big[\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, \theta_m, m\big]\Big], \qquad (13)$$

and C_T denotes the covariance with respect to the posterior. It should be stressed that these estimates of the mean and the covariance of the predictive distribution do not rely on the assumption that the latter distribution is normal.[15]

The normal approximation of the predictive likelihood for the observables $\mathcal{Y}_{s,T,h}^o$ can be computed from sample estimates of the moments in (12) and (13). The mean and covariance matrix of $\mathcal{Y}_{s,T,h}$ are determined by selecting the proper elements of (12) and (13), respectively. Next, notice that

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

with $\bar{n} = \sum_{i=1}^h n_i$, where

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

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

and $\epsilon_{s,T,h}^o(m)$ is the vector of prediction errors for the realizations $\mathcal{Y}_{s,T,h}^o$. The forecast uncertainty term is given by $D_{s,T,h}(m)$ in (15), while $Q_{s,T,h}(m)$ in (16) gives the impact of the quadratic standardized forecast errors on the normal approximation of the log predictive likelihood. This decomposition is of interest when the difference between the normal

[15] It is also worth noting that, according to equation (13), the covariance matrix of the predictive density is decomposed into two terms, where the first term on the right hand side reflects shock uncertainty over the forecast horizon (as well as uncertainty about unobserved variables up to period T) and the second term parameter uncertainty; see Adolfson, Lindé and Villani (2007). Geweke and Amisano (2014) more generally refer to the first term as the intrinsic variance of $\mathcal{Y}_{T,h}$ and the second term as the extrinsic variance.

approximation and the MC estimator is small, and it may then reveal whether forecast uncertainty (15) or forecast errors (16) is responsible for the ranking of models.

The log predictive scores estimated with the normal approximation are displayed in the lower part of Figure 1. It is noteworthy how similar they look when compared with the log predictive scores computed with the MC estimator in the upper part of the figure. The small numerical differences between the log scores using the MC estimator in (7) and the normal approximation in (14) for all models, forecast horizons, and selections of variables over the forecast sample are documented in Table E.1 in the online appendix.[16]

Since the normal approximation overall provides a good approximation of the MC estimator of the predictive likelihood, the decomposition in equations (14)–(16) is utilized to assess if the ranking of the models is driven by forecast uncertainty or by forecast errors. To this end, the contribution of the forecast uncertainty term to the recursive estimates of the average log predictive score are depicted in Panel A of Figure 4. Analogously, the quadratic standardized forecast error part of the recursive average log predictive scores are displayed in Panel B of Figure 4.

Starting with the forecast uncertainty term in Panel A we find that for all depicted models and forecast horizons this term is weakly upward sloping over the forecast sample and that the slope is roughly equal across the four models. This indicates that overall forecast uncertainty is slowly decreasing as data points are added to the historical sample. In general, the values for the BVAR model are roughly a few log-units higher in each period than for the second group of models, given by the DSGE-VAR and the NAWM.

[16] The differences between the MC estimator and the normal approximation of the log predictive score for the NAWM and the DSGE-VAR are positive for all forecast horizons and variable selections. In Table E.1 of the online appendix we also report, within parentheses, the differences between the two estimators when the quarters 2008Q4 and 2009Q1 are excluded from the calculations. As can be seen from the table these two quarters are particularly detrimental for the overall approximation errors for the BVAR model, but they also have a notable impact on approximation errors for the NAWM and the DSGE-VAR. In view of our findings above regarding the onset of the Great Recession, these results suggest that the accuracy of the normal approximation suffers when the value of the predictive likelihood is very low. Furthermore, a large fraction of the differences are positive suggesting that the normal approximation tends to be downwardly biased relative to the MC estimator (and the analytically determined value for the random walk model).

Overall, the ranking of models based on the forecast uncertainty term is constant across time periods, variable selections, and forecast horizons, with the exception of the small selection for the NAWM and the random walk model.

Turning to the quadratic standardized forecast error term in Panel B, it can be seen that the time variation of the recursively estimated average log predictive score is primarily due to the forecast errors. Second, the ranking of the models is to some extent reversed, particularly with the BVAR having much larger quadratic standardized forecast errors than the other models. 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 when the uncertainty linked to the forecast is smaller; see equation (16). Nevertheless, when compared with the forecast uncertainty term in Panel A the differences between the models are generally smaller for the forecast error term. Consequently, the model ranking based on the log predictive score is primarily determined by the second moments of the predictive distribution in this application. However, at the onset of the Great Recession the h-quarters-ahead forecast errors of the BVAR for the small selection with $h \geq 4$ are so severe that the model loses out to the two DSGE-based models.

4. Summary and Conclusions

This paper discusses how the predictive likelihood can be computed, by means of marginalization, for any subset of the observable variables in linear Gaussian discrete-time state-space models estimated with Bayesian methods. This approach is applied in an extension of the CCW study for euro area data and compares the results for the NAWM, a DSGE-VAR model with the NAWM as prior, a BVAR based on the large BVAR methodology of Bańbura et al. (2010), and a multivariate random walk model. The DSGE-VAR model was not included in CCW and is used to relax the possibly misspecified cross-equation restrictions of the NAWM, while the random walk model is an extension of the random walk model in CCW to a Bayesian framework. In addition, the forecast sample is extended by five years to include the Great Recession and ends in 2011Q4.

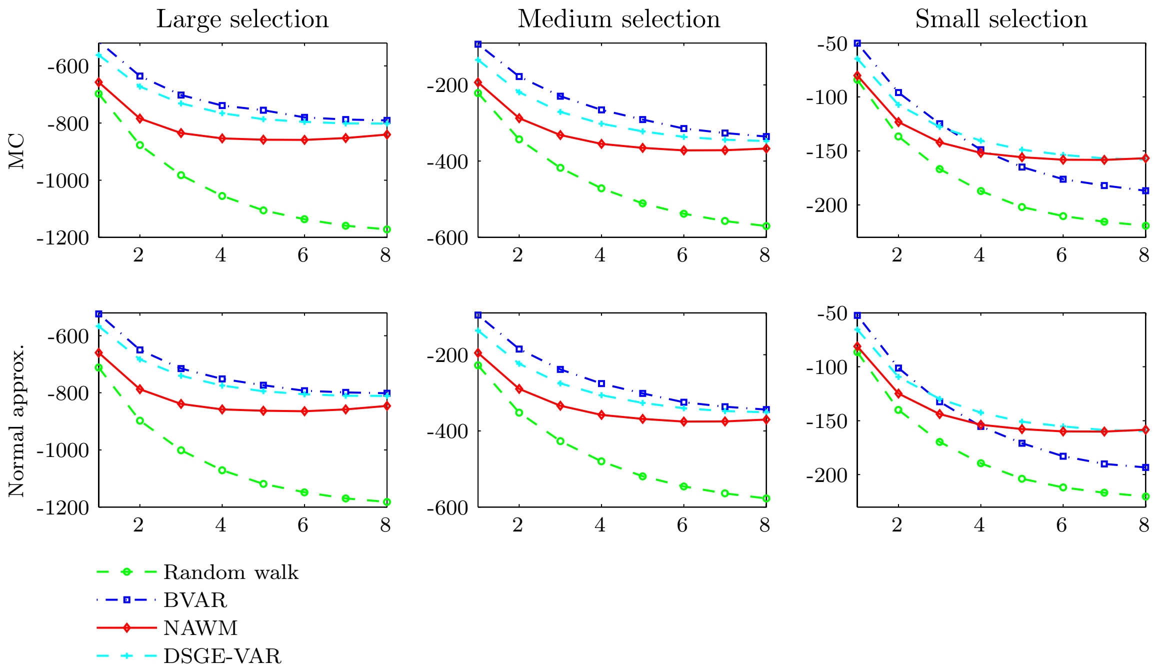
In terms of model ranking, the log predictive score (the sum of the log predictive likelihood over the forecast sample 1999Q1–2011Q4) typically favors the BVAR model, with the DSGE-VAR model improving somewhat on the density forecasts of the NAWM, especially at the shorter horizons. The random walk model, on the other hand, is only competitive with the NAWM at the one-step-ahead horizon, especially for the small variable selection with real GDP growth, GDP deflator inflation, and the short-term nominal interest rate.

It is noteworthy that for the longer-term forecasts and the small selection, the BVAR not only looses its first rank to the DSGE-VAR at the onset of the Great Recession in 2008Q4, but also the second rank to the NAWM. The main reason for this appears to be the deterioration in the BVAR density forecasts of real GDP growth, compared with those of the NAWM and the DSGE-VAR. In other words, the "more structural" models seem to cope better with the substantial loss in output growth observed during the Great Recession than the reduced form BVAR model.

In the empirical application, the Monte Carlo (MC) integration-based estimator of the marginalized predictive likelihood is compared with a normal approximation, constructed from the mean vector and the covariance matrix of the predictive distribution. We find that the assumption of a normal predictive density provides a good approximation of the predictive likelihood when examining the density forecasts of the four models. While the MC estimator is numerically reliable in our application, alternative estimators may be preferable in other density forecasting situations. For example, one may consider harmonic mean estimators of the marginalized predictive likelihood, cross entropy, or bridge sampling methods. Evaluating these options is beyond the scope of this paper, but it ought to be addressed in future research.

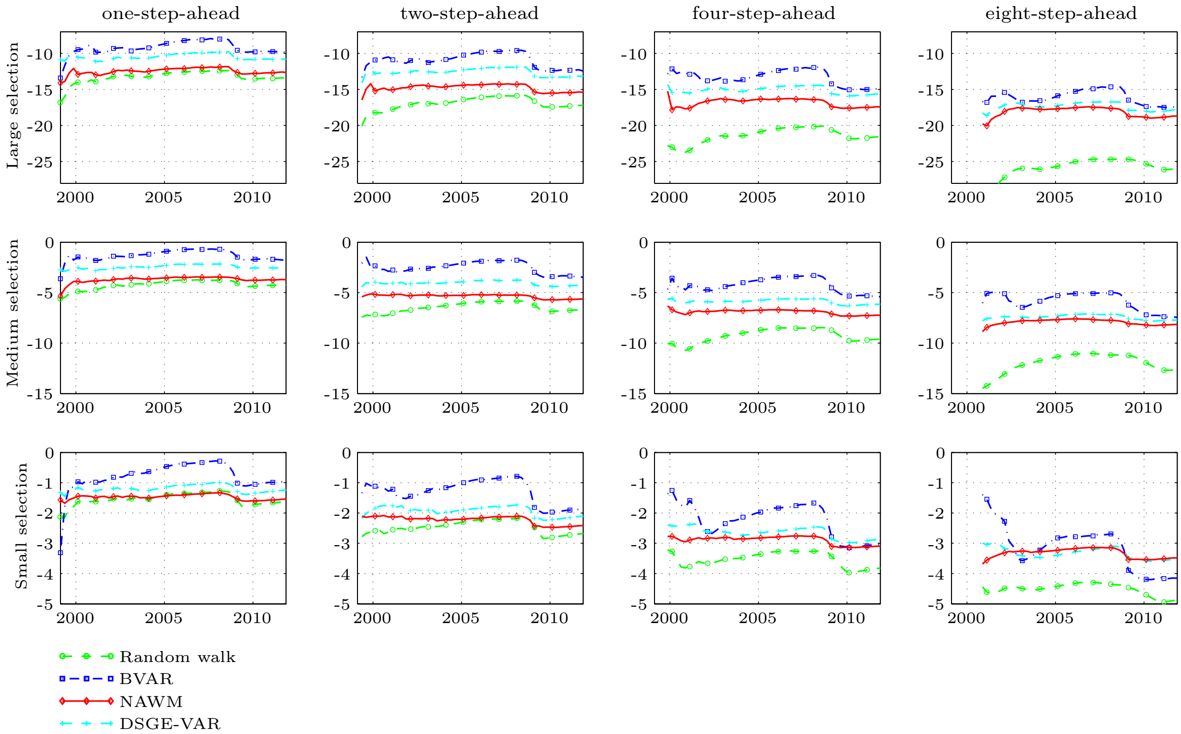
Although we have only considered linear Gaussian models that can be written in state-space form, this already covers a large number of the models frequently used in applied macroeconomics. The basic idea that has been presented for computing the marginalized conditional likelihood through a Kalman filter can, in principle, be extended to nonlinear and nonnormal models. For such models, the marginalized conditional likelihood may be estimated with a suitable particle filter; see, e.g., Giordani, Pitt and Kohn (2011) for a survey on filtering in state-space models. Whether or not this leads to a reliable approach for computing the marginalized predictive likelihood in such models, however, is an open and important question for future research.

FIGURE 1: Log predictive scores using the MC estimator of the predictive likelihood and the normal approximation for the entire sample 1999Q1–2011Q4.



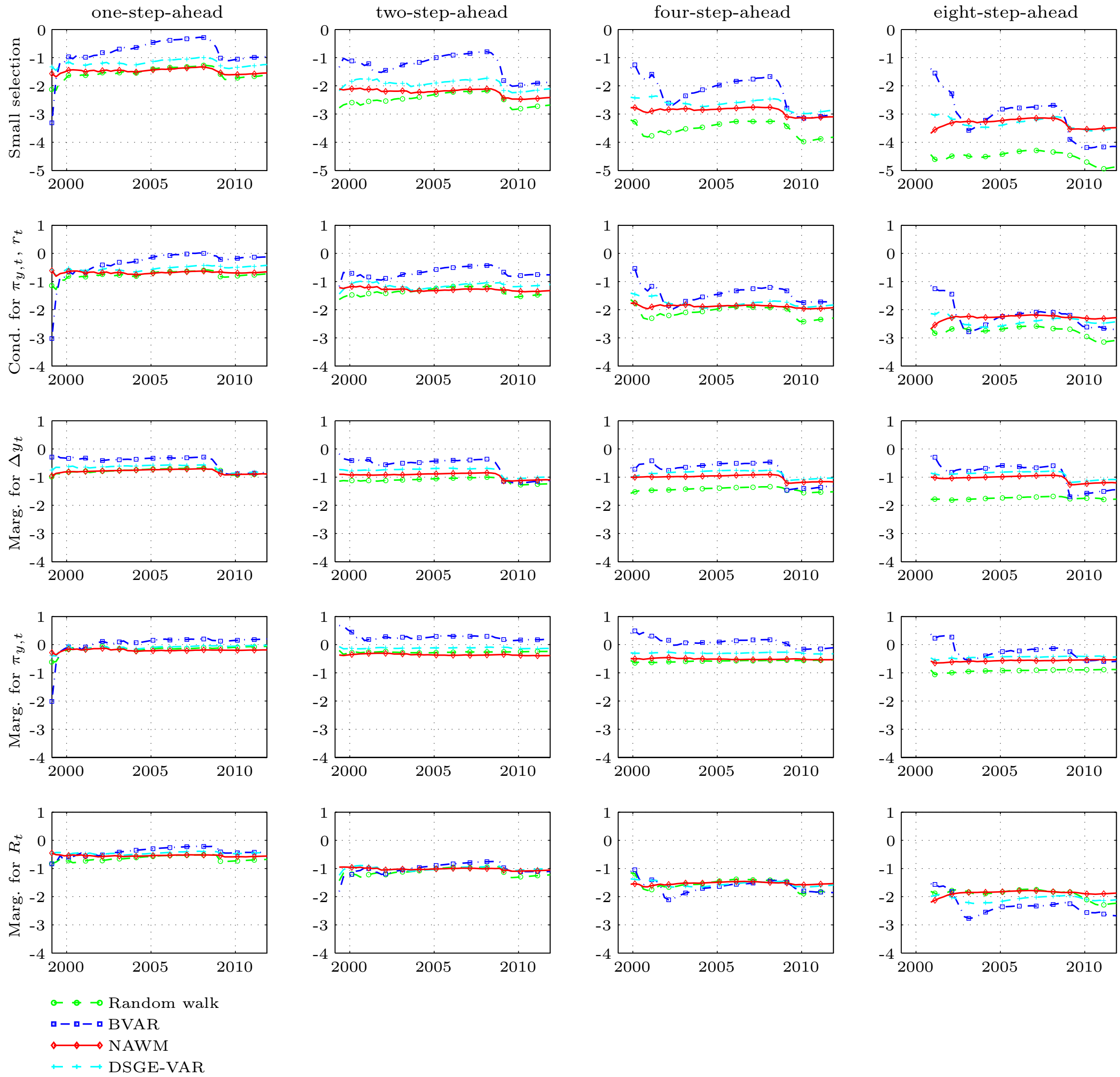
Note: The estimated log predictive score is given by equation (10) with $N_h = T_h = 53 - h$, the maximum number of time periods that the predictive likelihood can be evaluated for h-step-ahead density forecasts, and $h = 1, \ldots, 8$. The log predictive likelihood of the random walk model is calculated with its analytical expression in the MC estimator part of the figure.

FIGURE 2: Recursive estimates of the average log predictive score using the MC estimator for the sample 1999Q1–2011Q4.



Note: The recursively estimated average log predictive score is given by equation (11), with $N_h = 1, \ldots, T_h$, where $T_h = 53 - h$ for the forecast sample 1999Q1–2011Q4, while $h = 1, \ldots, 8$. The log predictive likelihood of the random walk model is calculated with its analytical expression.

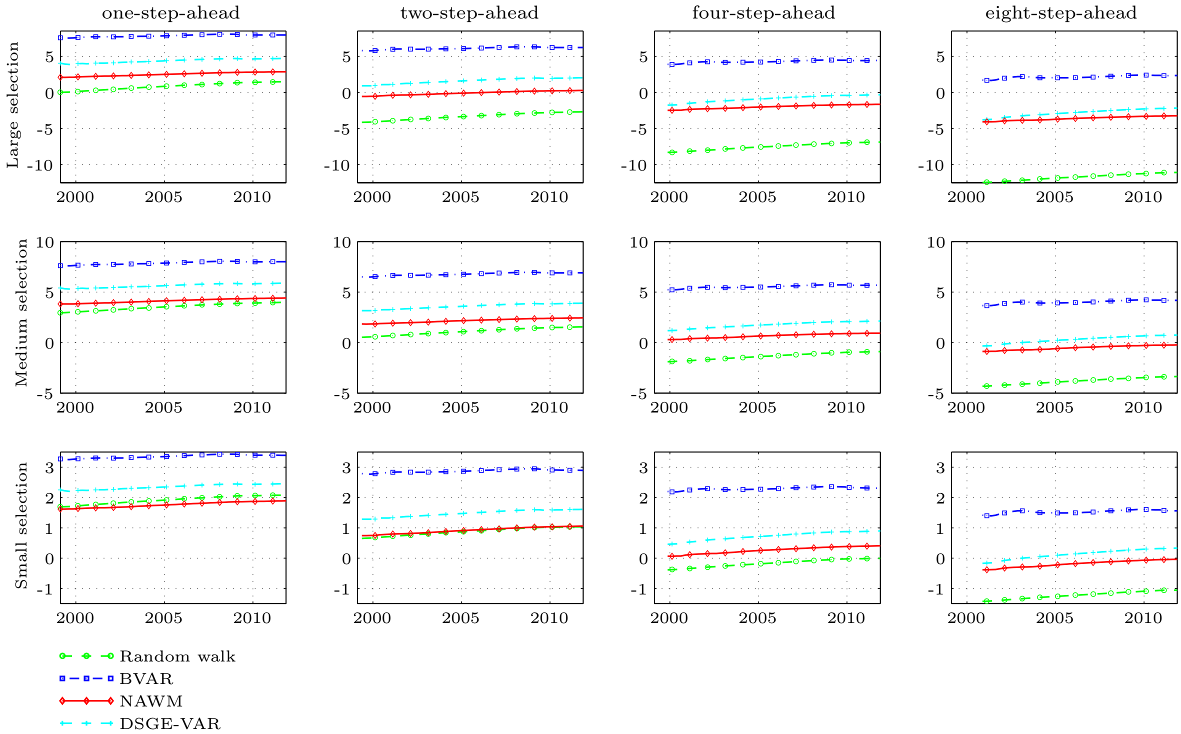
FIGURE 3: Recursive estimates of a decomposition of the average log predictive score for the small selection.



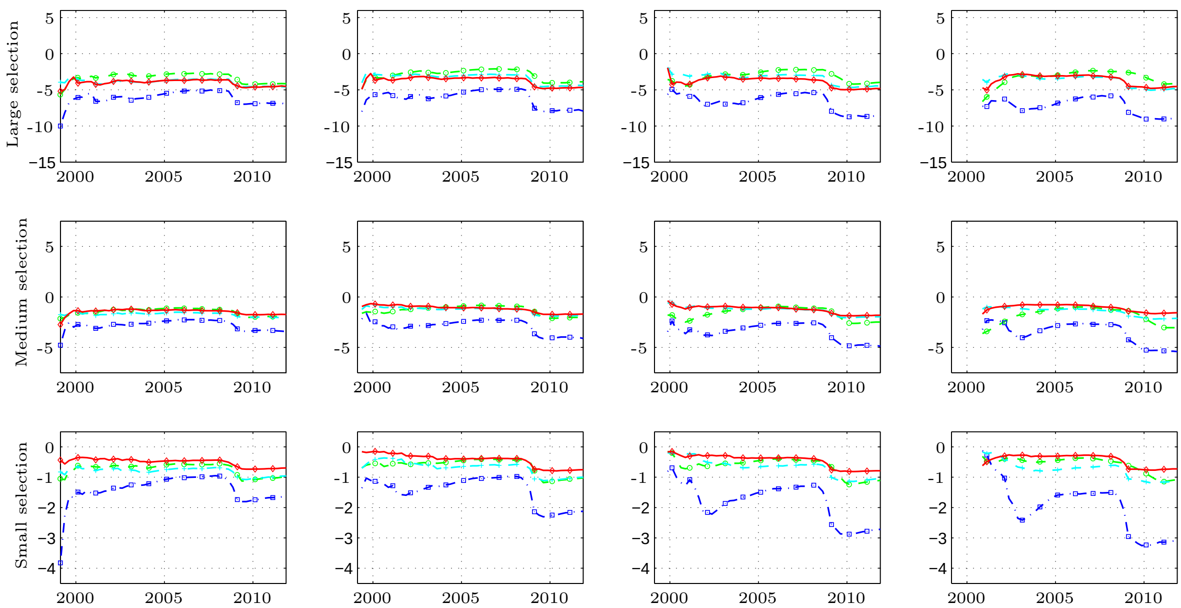
Note: The predictive likelihood of the small selection is decomposed into (i) the conditional predictive likelihood of GDP deflator inflation and the short-term nominal interest rate conditional on real GDP growth, and (ii) the marginal predictive likelihood of real GDP growth. For completeness, recursive estimates of the average log predictive score are also shown for GDP deflator inflation and the short-term nominal interest rate. Additional details are given in the note of Figure 2.

FIGURE 4: Recursive estimates of the decomposition of the average log predictive score into the forecast uncertainty term and the quadratic standardized forecast error term.

A. FORECAST UNCERTAINTY TERM



B. QUADRATIC STANDARIZED FORECAST ERROR TERM



Note: The forecast uncertainty and the quadratic standardized forecast error terms of the average log predictive score are computed from the normal approximation of the log predictive likelihood; see equations (15) and (16), respectively. Additional details are given in the note of Figure 2.

References

Adolfson M, Laséen S, Lindé J, Villani M. 2008. Evaluating an Estimated New Keynesian Small Open Economy Model. *Journal of Economic Dynamics and Control* **32**: 2690–2721. DOI: 10.1016/j.jedc.2007.09.012.

Adolfson M, Lindé J, Villani M. 2007. Forecasting Performance of an Open Economy DSGE Model. *Econometric Reviews* **26**: 289–328. DOI: 10.1080/07474930701220543.

An S, Schorfheide F. 2007. Bayesian Analysis of DSGE Models. *Econometric Reviews* **26**: 113–172. DOI: 10.1080/07474930701220071.

Andersson MK, Karlsson S. 2008. Bayesian Forecast Combinations for VAR Models. In Chib S, Koop G, Griffith W, Terrell D (eds.) *Bayesian Econometrics*. Bingley: Emerald Group Publishing, 501–524. Volume 23 of *Advances of Econometrics*.

Bańbura M, Giannone D, Reichlin L. 2010. Large Bayesian Vector Auto Regressions. *Journal of Applied Econometrics* **25**: 71–92. DOI: 10.1002/jae.1137.

Canova F. 2014. Bridging DSGE Models and the Raw Data. *Journal of Monetary Economics* **67**: 1–15. DOI: 10.1016/j.jmoneco.2014.06.003.

Chan JCC, Eisenstat E. 2015. Marginal Likelihood Estimation with the Cross-Entropy Method. *Econometric Reviews* **34**: 256–285. DOI: 10.1080/07474938.2014.944474.

Christoffel K, Coenen G, 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, Warne A. 2011. Forecasting with DSGE Models. In Clements MP, Hendry DF (eds.) *The Oxford Handbook of Economic Forecasting*. New York: Oxford University Press, 89–127.

Del Negro M, Hasegawa RB, Schorfheide F. 2014. Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance. NBER Working Paper No. 20575.

Del Negro M, Schorfheide F. 2004. Priors from General Equilibrium Models. *International Economic Review* **45**: 643–673. DOI: 10.1111/j.1468-2354.2004.00139.x.

Del Negro M, Schorfheide F. 2013. DSGE Model-Based Forecasting. In Elliott G, Timmermann A (eds.) *Handbook of Economic Forecasting*, volume 2. Amsterdam: North Holland, 57–140.

Del Negro M, Schorfheide F, Smets F, Wouters R. 2007. On the Fit of New-Keynesian Models. *Journal of Business & Economic Statistics* **25**: 123–143. DOI: 10.1198/073500107000000016.

Durbin J, Koopman SJ. 2012. *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press, 2nd edition.

Geweke J. 2005. *Contemporary Bayesian Econometrics and Statistics.* Hoboken: John Wiley.

Geweke J, Amisano G. 2010. Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. *International Journal of Forecasting* **26**: 216–230. DOI: 10.1016/j.ijforecast.2009.10.007.

Geweke J, Amisano G. 2014. Analysis of Variance for Bayesian Inference. *Econometric Reviews* **33**: 270–288. DOI: 10.1080/07474938.2013.807182.

Giannone D, Lenza M, Primiceri GE. 2015. Prior Selection for Vector Autoregressions. *The Review of Economics and Statistics* **97**: 436–451. DOI: 10.1162/REST_a_00483.

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

Gneiting T, Raftery AE. 2007. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association* **102**: 359–378. DOI: 10.1198/016214506000001437.

Good IJ. 1952. Rational Decisions. *Journal of the Royal Statistical Society Series B* **14**: 107–114.

Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Karlsson S. 2013. Forecasting with Bayesian Vector Autoregressions. In Elliott G, Timmermann A (eds.) *Handbook of Economic Forecasting*, volume 2. Amsterdam: North Holland, 791–897.

Litterman RB. 1986. Forecasting with Bayesian Vector Autoregressions — Five Years of Experience. *Journal of Business & Economic Statistics* **4**: 25–38. DOI: 10.1080/07350015.1986.10509485.

Sims CA. 1980. Macroeconomics and Reality. *Econometrica* **48**: 1–48.

Sims CA, Waggoner DF, Zha T. 2008. Methods for Inference in Large Multiple-Equation Markov-Switching Models. *Journal of Econometrics* **146**: 255–274. DOI: 10.1016/j.econom.2008.08.023.

Smets F, Warne A, Wouters R. 2014. Professional Forecasters and Real-Time Forecasting with a DSGE Model. *International Journal of Forecasting* **30**: 981–995. DOI: 10.1016/j.ijforecast.2014.03.018.

Smets F, Wouters R. 2003. An Estimated Stochastic Dynamic General Equilibrium Model for the Euro Area. *Journal of the European Economic Association* **1**: 1123–1175. DOI: 10.1162/154247603770383415.

Tierney L. 1994. Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics* **22**: 1701–1728. With discussion, p. 1728–1762.

Warne A, Coenen G, Christoffel K. 2013. Predictive Likelihood Comparisons with DSGE and DSGE-VAR Models. ECB Working Paper Series No. 1536.

Wolters MH. 2015. Evaluating Point and Density Forecasts of DSGE Models. *Journal of Applied Econometrics* **30**: 74–96. DOI: 10.1002/jae.2363.