CFS Working Paper Series

No. 478

Anders Warne, Günter Coenen, Kai Christoffel

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

Center for Financial Studies | Goethe University Frankfurt | Grüneburgplatz 1 | 60323 Frankfurt am Main | Germany
Phone +49 69 798 30050 | Fax +49 69 798 30077 | info@ifk-cfs.de | www.ifk-cfs.de

The CFS Working Paper Series

presents ongoing research on selected topics in the fields of money, banking and finance. The papers are circulated to encourage discussion and comment. Any opinions expressed in CFS Working Papers are those of the author(s) and not of the CFS.

The Center for Financial Studies, located in Goethe University Frankfurt's House of Finance, conducts independent and internationally oriented research in important areas of Finance. It serves as a forum for dialogue between academia, policy-making institutions and the financial industry. It offers a platform for top-level fundamental research as well as applied research relevant for the financial sector in Europe. CFS is funded by the non-profit-organization Gesellschaft für Kapitalmarktforschung e.V. (GfK). Established in 1967 and closely affiliated with the University of Frankfurt, it provides a strong link between the financial community and academia. GfK members comprise major players in Germany's financial industry. The funding institutions do not give prior review to CFS publications, nor do they necessarily share the views expressed therein.

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*

June 27, 2014

Abstract: The predictive likelihood is of particular relevance in a Bayesian setting when the purpose is to rank models in a forecast comparison exercise. This paper discusses how the predictive likelihood can be estimated for any subset of the observable variables in linear Gaussian state-space models with Bayesian methods, and proposes to utilize a missing observations consistent Kalman filter in the process of achieving this objective. As an empirical application, we analyze euro area data and compare the density forecast performance of a DSGE model to DSGE-VARs and reduced-form linear Gaussian models.

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

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

1. Introduction

In Bayesian analysis of (discrete) time series, the predictive likelihood can be employed to compare forecast accuracy across models. A problem occurs when the models included in the exercise predict different observables, but where some of these variables are shared by all models. For example, suppose there are two models to compare, where the first can predict the future values of the observables x_1, x_2, and x_3, while the second can predict the future values x_1 and x_4. To compare the forecast performance of these two models with the predictive likelihood requires that the likelihoods can be marginalized with respect to the non-shared variables, i.e. that the future values of x_2 and x_3 are integrated out from the predictive likelihood of the first model, while x_4 is integrated out for the second model.

The same basic marginalization problem arises when a forecaster or policy maker is concerned with a limited number of the variables that some multivariate models can predict. A special case is the forecast comparison of a single variable, such as inflation, across a set of models where

Note: This paper has previously been circulated with the title "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 on 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.

some may be univariate and others multivariate. Although point forecast accuracy statistics based on, e.g., mean squared forecast errors may reveal interesting aspects of how well the models involved can predict the shared variables, such statistics are not necessarily well suited for model selection (or ranking) since they neglect higher moments of the predictive distribution. The predictive likelihood, however, evaluates this distribution at the realization of the forecasted variables and is therefore a natural tool for addressing such matters, with the log predictive score as an extension from single-period comparisons to a multi-period setting.

The marginalization problem may be solved through textbook results when the joint predictive likelihood of a model has a known distributional form, for instance a Student t. However, models with a known predictive distribution are rare and in the typical case 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 times the posterior density with respect to the parameters. Here, by conditional likelihood we refer to the predictive likelihood *conditional* on a value for the parameters. If this likelihood is based on a distribution where marginalization can be handled analytically, then the marginalization problem of the predictive likelihood can be solved at this stage (see, e.g., Andersson and Karlsson, 2008, Karlsson, 2013, or Geweke and Amisano, 2010). What thereafter remains to be done is to integrate out the dependence of the parameters, a topic closely related to the estimation of the marginal likelihood; see, e.g., Geweke (2005).

In this paper we suggest a convenient and simple approach to marginalizing the conditional likelihood in linear Gaussian discrete-time state-space models, namely, to use a missing observations consistent Kalman filter. Since this approach builds up the marginalized parts of only the relevant arrays recursively, it is computationally simpler than first calculating the mean and the covariance matrix of the joint conditional distribution and thereafter reducing these.

As an empirical application, we use the suggested approach to enhance the forecast comparison exercise in Christoffel, Coenen, and Warne (2011), henceforth CCW. They review forecasting with dynamic stochastic general equilibrium (DSGE) models, using the New Area-Wide Model (NAWM)—an open-economy model of the euro area—as an example; see Christoffel, Coenen, and Warne (2008) for details about the NAWM. CCW's 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) observable variables that are endogenously determined in the NAWM. 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 marginalized conditional likelihood with respect to the posterior draws.

The remainder of the paper is organized as follows. Section 2 introduces notation and presents concepts related to the predictive likelihood. Section 3 focuses on linear state-space models with Gaussian innovations and shows how the conditional likelihood can be marginalized via a Kalman filter that takes missing observations into account. Section 4 presents the empirical density forecast comparison exercise, while Section 5 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 $\mathcal{Y}_T = \{y_1, y_2, \dots, y_T\}$ 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 of $\mathcal{Y}_T$ 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

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

The marginal likelihood is a standard measure to compare how well different models fit the data in Bayesian analysis and for each model m it is a joint assessment of how well the prior and likelihood agrees with the observations.

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

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

The joint predictive likelihood of model m is given by the predictive density in (2) evaluated at the observed values $\mathcal{Y}_{T,h}^o = \{y_{T+1}^o, \dots, y_{T+h}^o\}$.

Suppose that we are only interested in forecasting a subset of the variables $\mathcal{Y}_{T,h}$, denoted by $\mathcal{Y}_{s,T,h} = \{y_{s,T+1}, \dots, y_{s,T+h}\}$, where $y_{s,t} = S_t' y_t$, S_t is an $n \times n_{s,t}$ selection matrix with $n_{s,t} \in \{0, 1, \dots, n\}$ for $t = T+1, \dots, T+h$. Furthermore, let $y_{-s,t}$ denote the remaining variables in y_t, i.e. $y_{-s,t} = S_{t,\perp}' y_t$ with $S_{t,\perp}' S_t = 0$, while $S_t' S_t = I_{n_{s,t}}$ and $S_{t,\perp}' S_{t,\perp} = I_{n-n_{s,t}}$.

This general case covers the possibility that we are concerned with forecasting a sequence of future values of a particular variable, such as inflation, from period $T+1$ to $T+h$, as well as the marginal h-step-ahead forecast of some or all of the variables in the vector y_{T+h}. The marginalized predictive density can be expressed as

$$p(\mathcal{Y}_{s,T,h}|\mathcal{Y}_T^o, m) = \int_{\Theta_m} p(\mathcal{Y}_{s,T,h}|\mathcal{Y}_T^o, \theta_m, m)p(\theta_m|\mathcal{Y}_T^o, m)d\theta_m, \tag{3}$$

where

$$p(\mathcal{Y}_{s,T,h}|\mathcal{Y}_T^o, \theta_m, m) = \int \cdots \int p(\mathcal{Y}_{T,h}|\mathcal{Y}_T^o, \theta_m, m)dy_{-s,T+1} \cdots dy_{-s,T+h}. \tag{4}$$

The marginalized predictive likelihood is given by (3) when evaluated at the observed values $\mathcal{Y}^o_{s,T,h}$, while equation (4) evaluated at $\mathcal{Y}^o_{s,T,h}$ is called the marginalized *conditional* likelihood.

The order in which the parameters θ_m and the variables $y_{-s,t}$ are integrated out can of course be reversed when determining the left hand side of (3). However, the precise density function which determines the conditional likelihood is typically an integral part of the model assumptions in a Bayesian setting, suggesting that this function may be "easier" to operate on when attempting to integrate out the dependence on the variables $y_{-s,t}$ than the joint predictive density, which will generally not be of a known type.

The marginalized predictive likelihood can alternatively be expressed as the ratio of two marginal likelihoods:

$$p\bigl(\mathcal{Y}^o_{s,T,h}\big|\mathcal{Y}^o_T, m\bigr) = \frac{p\bigl(\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T\big|m\bigr)}{p\bigl(\mathcal{Y}^o_T\big|m\bigr)}. \tag{5}$$

The denominator is here given by equation (1), while the numerator is

$$p\bigl(\mathcal{Y}^o_{s,T,h}, \mathcal{Y}^o_T\big|m\bigr) = \int_{\Theta_m} p\bigl(\mathcal{Y}^o_{s,T,h}\big|\mathcal{Y}^o_T, \theta_m, m\bigr)p\bigl(\mathcal{Y}^o_T\big|\theta_m, m\bigr)p\bigl(\theta_m\big|m\bigr)d\theta_m. \tag{6}$$

Again we see that the computation of the marginalized predictive likelihood depends on being able to compute the marginalized conditional likelihood.

In the next Section we suggest an approach to marginalize the conditional likelihood in equation (4) for $\mathcal{Y}^o_{s,T,h}$ 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 (3) for $\mathcal{Y}^o_{s,T,h}$ generally depends on selecting an appropriate numerical method for integrating out the dependence on the parameters. This is closely related the problem of estimating the marginal likelihood; see, e.g., Geweke (2005).

3. Linear State-Space Models with Gaussian Innovations

3.1. Marginalizing the Conditional Likelihood

Standard linear time series models—including finite order 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 underlying parameter vector.[1]

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 equation

$$y_t = \mu + H'\xi_t + w_t, \quad t = 1, \dots, T. \tag{7}$$

[1] Sargent (1989) was among the first to recognize that linear rational expectations models can be expressed in state-space form.

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 VAR system:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1,\ldots,T. \tag{8}$$

The state shocks, η_t, are of dimension q and i.i.d. $N(0,I_q)$, 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 all uniquely determined by θ_m. Provided that $H'\xi_t$ is stationary, the vector μ is the population mean of y_t conditional on θ_m.

The system in (7) and (8) is a state-space model, where equation (7) gives the measurement or observation equation and (8) 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) and Durbin and Koopman (2012) for details.[2]

Suppose that 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 $\mathcal{Y}^o_{s,T,h}$. The log of the conditional likelihood can now be determined via a missing observations consistent Kalman filter. We here find for $h \geq 1$

$$\log p\big(\mathcal{Y}^o_{s,T,h}\big|\mathcal{Y}^o_T, \theta_m, m\big) = \sum_{i=1}^{h} \log p\big(y^o_{s,T+i}\big|\mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big), \tag{9}$$

where $\mathcal{Y}^o_{s,T,0}$ is empty by definition. If $n_{s,T+i} \geq 1$ the log-likelihood value at $T+i$ is

$$\begin{aligned}\log p\big(y^o_{s,T+i}\big|\mathcal{Y}^o_{s,T,i-1}, \mathcal{Y}^o_T, \theta_m, m\big) &= -\frac{n_{s,T+i}}{2}\log(2\pi) - \frac{1}{2}\log\big|\Sigma_{y_s,T+i|T+i-1}\big| \\ -\frac{1}{2}\big(y^o_{s,T+i} - y_{s,T+i|T+i-1}\big)' &\Sigma^{-1}_{y_s,T+i|T+i-1}\big(y^o_{s,T+i} - y_{s,T+i|T+i-1}\big),\end{aligned} \tag{10}$$

while it is zero when $n_{s,T+i} = 0$.

Let $\mu_{s,t} = S_t'\mu$, $H_{s,t} = HS_t$, and $R_{s,t} = S_t'RS_t$ when $n_{s,t} \geq 1$ for $t = T+1,\ldots,T+h$. The one-step-ahead forecasts of $y_{s,t+1}$ and the forecast error covariance matrix conditional on information available at time t for $t = T,\ldots,T+h-1$ are when $n_{s,t+1} \geq 1$ given by

$$\begin{aligned} y_{s,t+1|t} &= \mu_{s,t+1} + H'_{s,t+1}\xi_{t+1|t}, \\ \Sigma_{y_s,t+1|t} &= H'_{s,t+1}P_{t+1|t}H_{s,t+1} + R_{s,t+1}. \end{aligned}$$

If $n_{s,t} = 0$, the one-steap-ahead forecast of the state variables and the corresponding forecast error covariance matrix are

$$\begin{aligned} \xi_{t+1|t} &= F\xi_{t|t-1} = F\xi_{t|t}, \\ P_{t+1|t} &= FP_{t|t-1}F' + BB' = FP_{t|t}F' + BB', \end{aligned}$$

[2] It is straightforward to generalize the system in (7) and (8) to allow for, e.g., time variation of the matrices and apply the ideas presented in this paper.

while if $n_{s,t} \geq 1$ we instead obtain

$$\xi_{t+1|t} = F\xi_{t|t-1} + K_t\big(y^o_{s,t} - y_{s,t|t-1}\big) = F\xi_{t|t},$$
$$P_{t+1|t} = \big(F - K_t H'_{s,t}\big)P_{t|t-1}\big(F - K_t H'_{s,t}\big)' + K_t R_{s,t} K'_t + BB' = FP_{t|t}F' + BB',$$
$$K_t = FP_{t|t-1}H_{s,t}\Sigma^{-1}_{y_s,t|t-1}.$$

The equations for $\xi_{t+1|t}$ and $P_{t+1|t}$ are initialized by $\xi_{T|T}$ and $P_{T|T}$, respectively, obtained from the Kalman filter estimates of the state variables using $\mathcal{Y}^o_T$. Notice that the Kalman filter based approach generates a bottom-up evaluation of the marginalized conditional likelihood.

An interesting special case of the above filtering equations arises when we are concerned with the marginal h-step-ahead forecast of $y_{s,T+h}$. We here have that $n_{s,t} = 0$ for $t = T+1, \ldots, T+h-1$, while

$$y_{s,T+h|T} = \mu_{s,T+h} + H'_{s,T+h}F^h\xi_{T|T},$$
$$\Sigma_{y_s,T+h|T} = H'_{s,T+h}P_{T+h|T}H_{s,T+h} + R_{s,T+h},$$
$$P_{T+h|T} = FP_{T+h-1|T}F' + BB'.$$

Since the joint conditional likelihood for $\mathcal{Y}^o_{T,h}$ is normal, the marginalized conditional likelihood 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. As an alternative, it is therefore possible to follow a procedure for computing the marginalized conditional likelihood based on a setup similar to the one suggested by Schmitt-Grohé and Uribe (2010) for evaluating the log-likelihoood. This top-down approach to evaluating the marginalized conditional likelihood is expected to be much slower than the Kalman filter calculations when nh is large, especially when repeated many times.

3.2. 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, to integrate out the dependence on the parameters. It is assumed below that parameter draws from the posterior density based on an ergodic posterior simulator are available.

In the empirical application in Section 4 we apply a simple Monte Carlo integration method to estimate the predictive likelihood. Specifically, we use:

$$\hat{p}_{MC}\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, m\big) = \frac{1}{N}\sum_{j=1}^{N} p\big(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta^{(j)}_m, m\big), \tag{11}$$

where $\theta^{(j)}_m$ is a draw from the posterior distribution $p(\theta_m|\mathcal{Y}^o_T)$ for $j = 1, \ldots, N$. Under certain regularity conditions (Tierney, 1994), the right hand side of (11) converges almost surely to the expected value of $p(\mathcal{Y}^o_{s,T,h}|\mathcal{Y}^o_T, \theta_m, m)$ with respect to $p(\theta_m|\mathcal{Y}^o_T)$, i.e. to the predictive likelihood

$p(\mathcal{Y}_{s,T,h}^{o}|\mathcal{Y}_{T}^{o}, m)$. 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, 2013, Proposition 1).[3]

The MC estimator in (11) 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 questionable when one of these properties is not met. For instance, when $S_t = I_n$ for all t and h is sufficiently large, the situation resembles the case when the marginal likelihood is estimated by averaging the likelihood over the prior draws, and such an estimator is typically expected to be poor.

One may therefore consider alternative standard methods for calculating the marginal likelihood, such as harmonic mean estimators; see Gelfand and Dey (1994), the truncated normal version in Geweke (2005), or the extension to a truncated elliptical in Sims, Waggoner, and Zha (2008). Such estimators require two sets of posterior draws to be consistent. Namely, apart from the $\theta_m^{(j)}$ draws based on the data $\mathcal{Y}_T^o$, we would also need posterior draws $\theta_m^{(i)}$ using the data $(\mathcal{Y}_{s,T,h}^o, \mathcal{Y}_T^o)$. In other words, harmonic mean estimators of the marginalized predictive likelihood will be more time consuming than the MC estimator in (11). However, since the harmonic mean estimators take the data $\mathcal{Y}_{s,T,h}^o$ into account, they are less likely to be hampered by a large dimension of the set of predicted values or a long forecast horizon. Other methods, such as bridge sampling may also be considered; see, e.g., Meng and Wong (1996). To evaluate these options is beyond the scope of this paper, but is an important topic for future reasearch.

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

In this Section we compare marginalized h-step-ahead density forecasts for three subsets of the observables and across five linear Gaussian state-space models for euro area data using the approach discussed in Section 3. In Section 4.1 we discuss the models, a DSGE model, two DSGE-VAR models, a large BVAR model, as well as a multivariate random walk model. In Section 4.2, we present the forecast sample and summarize the empirical results of the exercise. The calculations below have been performed with the help of YADA, a Matlab program for Bayesian estimation and evaluation of DSGE and DSGE-VAR models; see Warne (2014) for details.

4.1. Empirical Models

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

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 methodological

[3] The numerical standard error of (11) can be calculated using the Newey and West (1987) estimator when the posterior draws are correlated.

advances 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 (2013), Smets, Warne, and Wouters (2014) and references therein.

With E_t being the rational expectations operator, a log-linearized DSGE model can be written:

$$A_{-1}\xi_{t-1} + A_0\xi_t + A_1 E_t\xi_{t+1} = D\eta_t, \quad t = 1, 2, \ldots, T, \tag{12}$$

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. 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 (12) exists at a particular value of these parameters,[4] we can express the model variables as the first order VAR system in (8).

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 NAWM with Bayesian methods. They 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

[4] See, e.g., Anderson (2010), Klein (2000), or Sims (2002) for methods of solving linear rational expectations models.

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. See Christoffel et al. (2008) for details on the NAWM's estimation properties.

4.1.2. DSGE-VAR Models 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 y_t can be written as:

$$y_t = \Phi_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \epsilon_t, \quad t = 1, \ldots, T, \tag{13}$$

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 for 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 (see, e.g., Del Negro and Schorfheide, 2011, or Karlsson, 2013) 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. A well-known example when the model includes additional parameters through the prior is a DSGE-VAR.

An early attempt to combine DSGE models with VARs is Ingram and Whiteman (1994), where the VAR parameters are expressed as a function of the DSGE model parameters. A prior for the DSGE model parameters then implies a prior for the VAR parameters through a first-order Taylor expansion of the mapping. This idea is considerably enriched by Del Negro and Schorfheide (2004), where the prior distribution of the VAR model parameters conditional on the DSGE model parameters is specified such that the conditional moments are determined through the implied first and second population moments of y_t 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 of 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 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.

The DSGE-VAR models with the NAWM as prior have been estimated by Warne, Coenen, and Christoffel (2013) with the random walk Metropolis algorithm subject to a Gaussian proposal density. They 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. Based on lag orders between one and four lags, the first approach selects the DSGE-VAR model with $(\lambda, p) = (2.5, 2)$, while for the second approach favors the model with $(\lambda, p) = (6, 4)$. For details and further discussions, see Warne et al. (2013).

4.1.3. VAR and Random Walk Models

We also consider a Bayesian VAR model for the same observable 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. Below, we employ the same large BVAR as in CCW, estimated using the methodology in Bańbura, Giannone, and Reichlin (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 also 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. $N(0, \Omega)$, 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 marginal h-step-ahead forecasts of $y_{s,T+h}$ the predictive density is given by a n_s-dimensional Student t-distribution with mean $y^o_{s,T}$, covariance matrix

$$\frac{h}{T-n-1} \sum_{t=1}^{T} \big(y^o_{s,t} - y^o_{s,t-1}\big)\big(y^o_{s,t} - y^o_{s,t-1}\big)',$$

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

4.2. 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 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

$$S_h(m) = \sum_{t=T}^{T+T_h-1} \log p(\mathcal{Y}_{s,T,h}|\mathcal{Y}_t^o, m), \quad h = 1, \ldots, h^*, \tag{14}$$

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 observations; 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 below we will evaluate the log predictive score in (14) 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. 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.

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 calculations in Section 3 can be adjusted to handle such transformations of the observables; see Warne (2014, 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.

The log predictive scores based on the MC estimator of the marginal h-step-ahead predictive likelihood are shown in Figure 2 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 MC 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.[5]

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. 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.

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.

[5] The estimation time of the predictive likelihood for the NAWM is negligable compared with posterior sampling using the random-walk Metropolis algorithm. Furthermore, the estimation time of the predictive likelihood for the DSGE-VAR models is more than halved relative to the NAWM time.

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 et al. (2013) for additional details on the results.

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 necessary when we know how to compute the conditional likelihood, it can nevertheless serve as a tool for enhancing our understanding of the results from a forecast comparison study. Moreover, it may be interesting to estimate the size of the errors from using a normal approximation relative to the MC estimator, when the latter is assumed to be accurate.

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 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 given by

$$E\bigl[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o,m\bigr] = E_T\Bigl[E\bigl[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o,\theta_m,m\bigr]\Bigr], \tag{15}$$

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

$$C\bigl[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o,m\bigr] = E_T\Bigl[C\bigl[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o,\theta_m,m\bigr]\Bigr] + C_T\Bigl[E\bigl[\mathcal{Y}_{T,h}\big|\mathcal{Y}_T^o,\theta_m,m\bigr]\Bigr], \tag{16}$$

and C_T denotes the covariance with respect to the posterior. Notice that the covariance matrix of the predictive density is decomposed into two terms, where the first term on the right hand side reflects residual uncertainty and the second term parameter uncertainty; see Adolfson et al. (2007b). Geweke and Amisano (2014) more generally refer to the first term as intrinsic variance of $\mathcal{Y}_{T,h}$ and the second term as extrinsic variance.

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 (15) and (16), respectively. Next, notice that

$$\log \hat{p}_N\bigl(\mathcal{Y}_{s,T,h}^o\big|\mathcal{Y}_T^o,m\bigr) = -\frac{d}{2}\log\bigl(2\pi\bigr) + D_{s,T,h}(m) + Q_{s,T,h}(m), \tag{17}$$

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

$$D_{s,T,h}(m) = -\frac{\log \left| C\left[\mathcal{Y}_{s,T,h} \middle| \mathcal{Y}_T^o, m\right] \right|}{2}, \tag{18}$$

$$Q_{s,T,h}(m) = -\frac{\epsilon^{o\prime}_{s,T,h}(m) C\left[\mathcal{Y}_{s,T,h} \middle| \mathcal{Y}_T^o, m\right]^{-1} \epsilon^{o}_{s,T,h}(m)}{2}, \tag{19}$$

and $\epsilon^o_{s,T,h}(m)$ the the vector of prediction errors for the realizations $\mathcal{Y}^o_{s,T,h}$. The forecast uncertainty term is given by $D_{s,T,h}(m)$, while $Q_{s,T,h}(m)$ 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 MC estimator is small, or the ranking of models is robust across these measures. For such cases, the decomposition in (17) may reveal whether forecast uncertainty (18) or forecast errors (19) is responsible for the ranking of models.

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 3. The most prominent feature is how similar these graphs are when compared to those in Figure 2. In fact, the MC 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 numerical differences in log predictive score between the MC estimator and the normal approximation for all models, forecast horizons, and selections of variables are documented in Table 1.

The differences between the MC 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 weak tendency of 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 aggregate 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 likelihood values are smaller. Further details are shown in Warne et al. (2013).

Since the normal approximation appears to provide a good approximation of the MC estimator of the predictive likelihood for the five models and the three variable selections, it may be of interest to utilize equations (17)–(19) to assess if the ranking of the models is driven by forecast uncertainty or by forecast errors. The forecast uncertainty term in equation (18) of the decomposition of the log predictive likelihood is depicted in Figure 4 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 (19) is displayed in Figure 5 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 4 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 4 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 5, it may be deduced that time variation of the log predictive likelihood is primarily 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 in Figure 4 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 application.

5. Summary and Conclusions

This paper discusses how the predictive likelihood can be computed for any subset of the observable variables in linear Gaussian discrete-time state-space models estimated with 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 is based on the assumption 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 considered 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 (Eklund and Karlsson, 2007), or when constructing optimal prediction pools (Geweke and Amisano, 2011, 2012) under a Bayesian approach. The paper suggests that the marginalized predictive likelihood for linear Gaussian discrete-time state-space models is computed via a missing observations consistent Kalman filter. This approach builds up the marginalized parts of only the relevant arrays recursively and is therefore simpler and faster than first calculating the mean and the covariance matrix of the joint predictive likelihood conditional on the parameters (conditional likelihood), and thereafter reducing these arrays to the entries for the set of variables of interest.

The missing observations consistent Kalman filter instead evaluates the conditional likelihood for the variables of interest only using simple recursions. Based on such a value for each posterior draw of the parameters, the paper considers simple Monte Carlo integration over the posterior draws to estimate of the marginalized predictive likelihood. This MC estimator is expected to work well in practise when the posterior draws cover well enough the parameter region where the conditional likelihood is large.

In the empirical application with five linear Gaussian models, the MC 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 application 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 MC estimator for all models except the random walk, whose predictive density is multivariate Student 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 MC estimator is small, then the analytical form of the normal density can be utilized to assess which feature of the predictive likelihood drives 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. In our empirical application, this decomposition suggests that the model ranking is primarily influenced by the forecast uncertainty term, while the forecast errors are mainly responsible for the volatility in the predictive likelihood.

We have also discussed some alternative estimators of the marginalized predictive likelihood which are likely to work better than the simple MC estimator when the number of predicted variables is large. These estimator also require that the conditional likelihood can be marginalized and, thus, benefit from the ideas presented in the paper. An interesting topic for future research is to examine such methods in more detail and thereby to learn more about, e.g., the limitations of the simple MC estimator we have used in this paper.

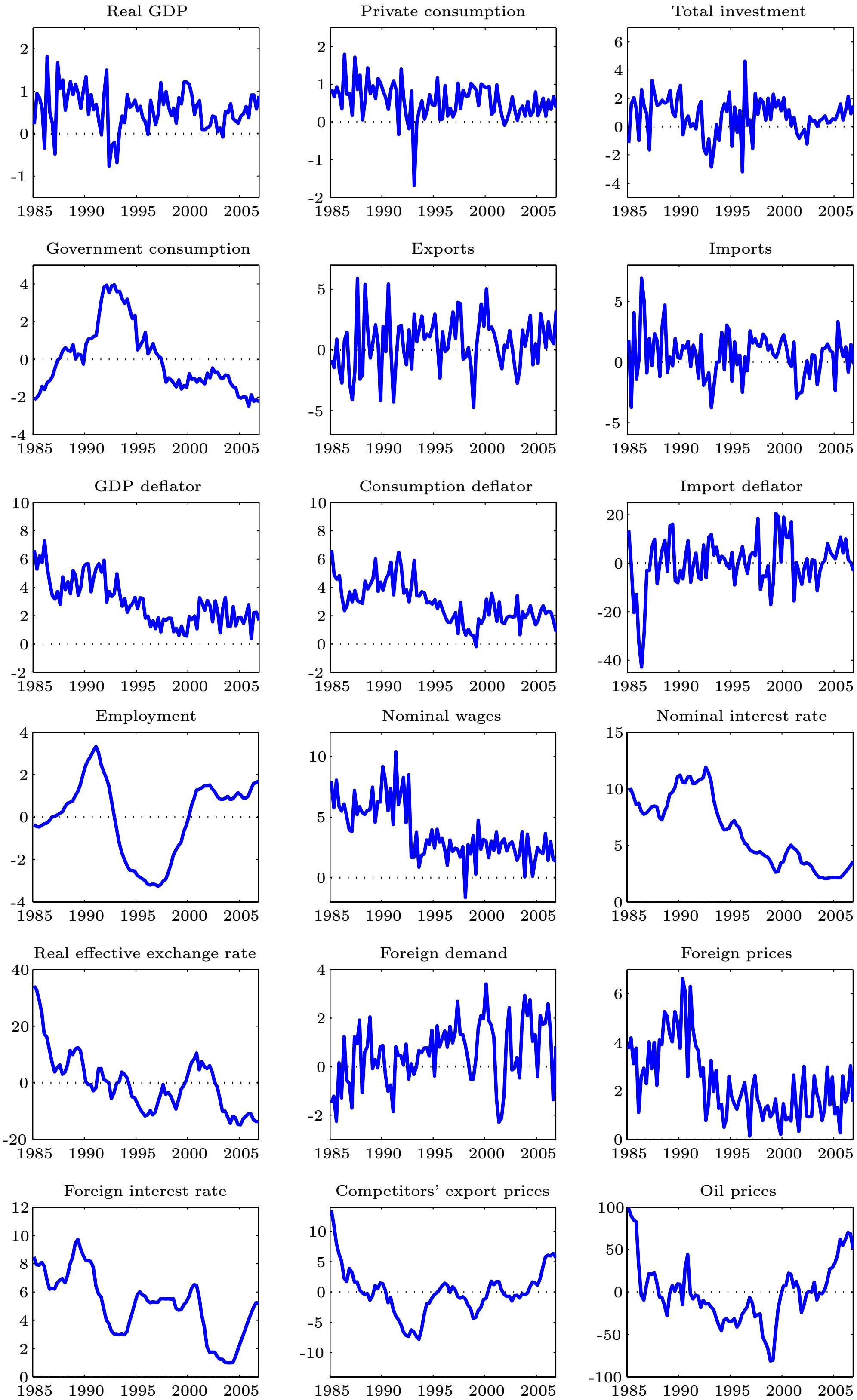
Although we have only discussed linear Gaussian models that can be written in state-space form, this already covers a large number of the models frequently used in, for example, applied macroeconomics. The basic idea that has been presented for computing the conditional likelihood through a missing observations consistent 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 missing observations consistent particle filter (sequential Monte Carlo); see, e.g., Giordani, Pitt, and Kohn (2011) for a survey on filtering in state-space models, or Durbin and Koopman (2012, Chapter 12) for an introduction to particle filtering. If this leads to a reliable approach for computing the marginalized conditional likelihood or not in such models, however, is an open and important question for future research.

TABLE 1: Difference between log predictive score using the MC 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

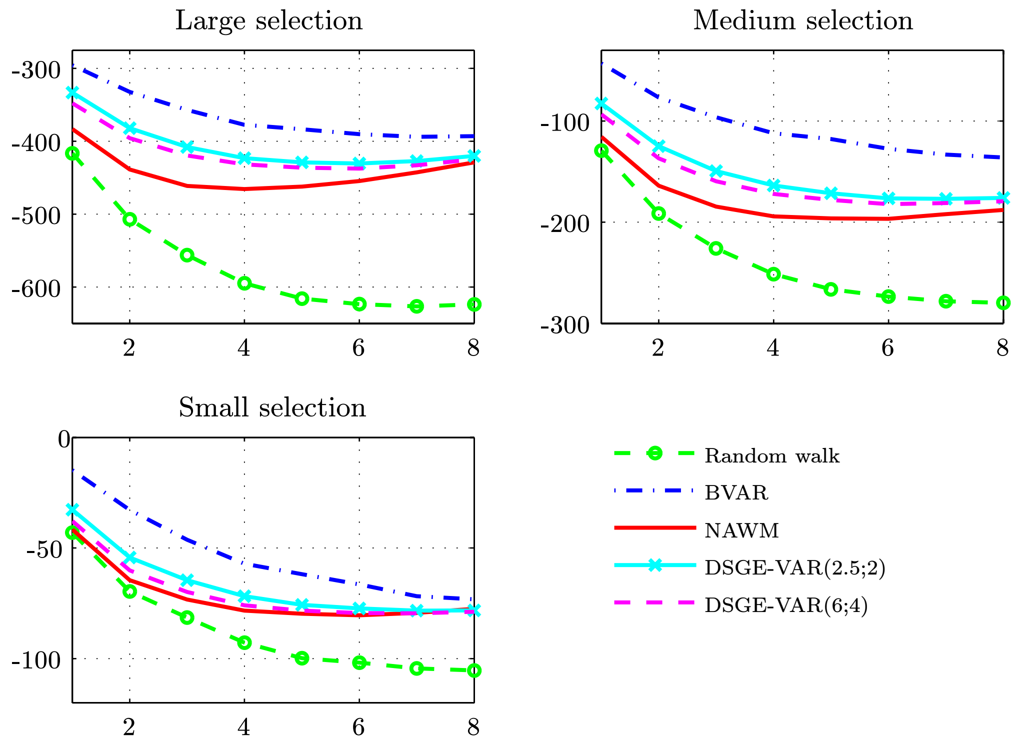
Note: The log predictive likelihood for the random walk model is calculated with its analytical expression; see the online appendix (Appendix A) for details. For the NAWM and the DSGE-VAR models, 10,000 posterior draws have been taken from the available 500,000 post burn-in draws for each time period. 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 observable 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: Log predictive scores using the MC estimator.



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

Figure 3: Log predictive scores using the normal approximation.

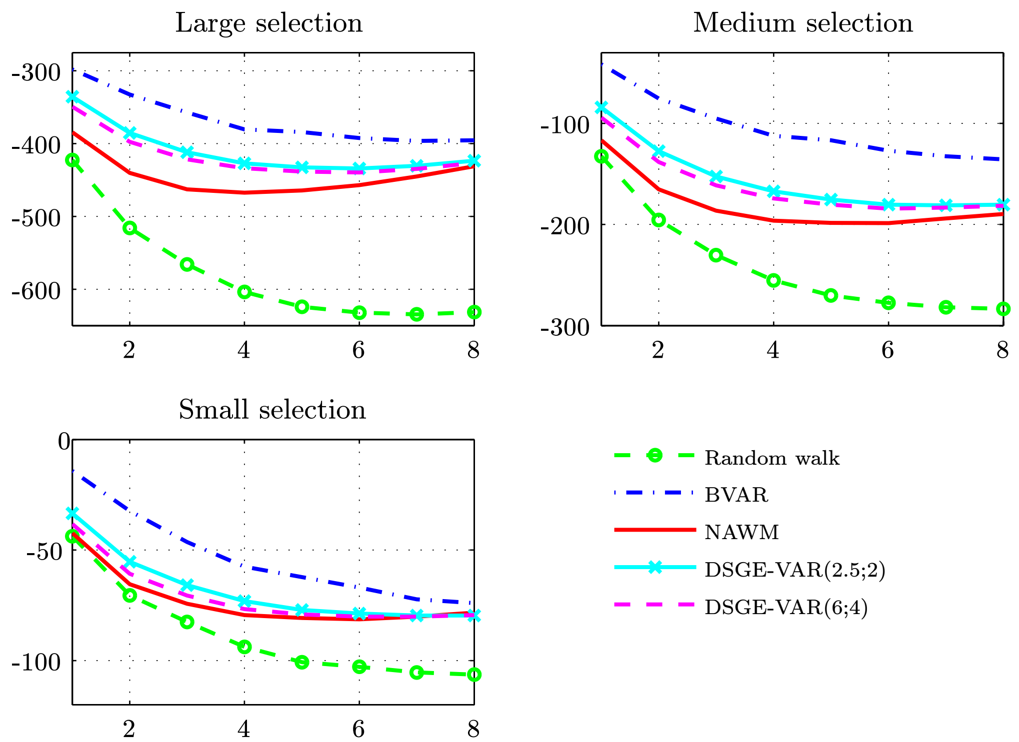


FIGURE 4: The evolution of the forecast uncertainty term of the normal density for the large selection of variables.

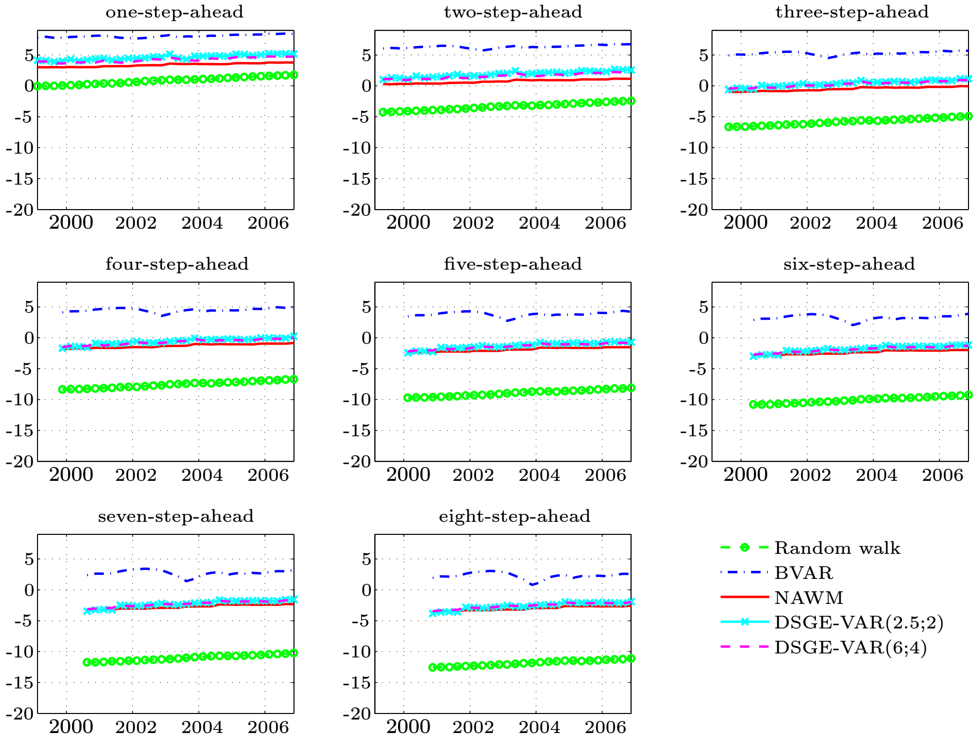
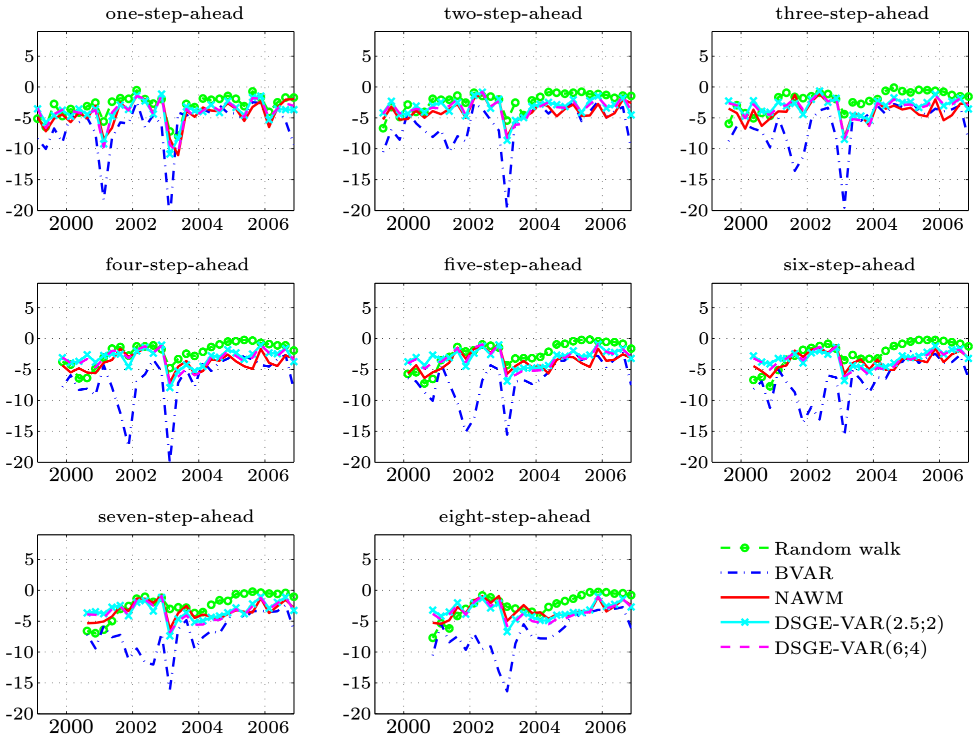


FIGURE 5: The evolution of the quadratic standardized forecast error term of the normal density for the large selection of variables.



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., 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.

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. (2011), "Bayesian Macroeconometrics," in J. Geweke, G. Koop, and H. K. van Dijk (Editors), *The Oxford Handbook of Bayesian Econometrics*, 293–389, Oxford University Press, New York.

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

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

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

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

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

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

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

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

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

Geweke, J. and Amisano, G. (2014), "Analysis of Variance for Bayesian Inference," *Econometric Reviews*, 33, 270–288.

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.

Harvey, A. C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

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. (2013), "Forecasting with Bayesian Vector Autoregressions," in G. Elliott and A. Timmermann (Editors), *Handbook of Economic Forecasting*, volume 2, 791–897, North Holland, Amsterdam.

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.

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.

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

Schmitt-Grohé, S. and Uribe, M. (2010), "Evaluating the Sample Likelihood of Linearized DSGE Models without the Use of the Kalman Filter," *Economics Letters*, 109, 142–143.

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., Warne, A., and Wouters, R. (2014), "Professional Forecasters and Real-Time Forecasting with a DSGE Model," *International Journal of Forecasting*, Forthcoming.

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.

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

Warne, A. (2014), "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," ECB Working Paper Series No. 1536.

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

CFS Working Paper Series

Recent Issues

All CFS Working Papers are available at *www.ifk-cfs.de*.

No.	Authors	Title
477	Markus Bibinger, Nikolaus Hautsch, Peter Malec, Markus Reiss	*Estimating the Spot Covariation of Asset Prices – Statistical Theory and Empirical Evidence*
476	Christina Bannier, Markus Wiemann	*Performance-Sensitive Debt – the Intertwined Effects of Performance Measurement and Pricing Grid Asymmetry*
475	Christina Bannier, Eberhard Feess, Natalie Packham	*Incentive Schemes, Private Information and the Double-Edged Role of Competition for Agents*
474	Steven Ongena, Ibolya Schindele, Dzsamila Vonnak	*In Lands of Foreign Currency Credit, Bank Lending Channels Run Through?*
473	Fabian Kindermann, Dirk Krueger	*High Marginal Tax Rates on the Top 1%?*
472	Gianmarco Ottaviano, João Paulo Pessoa, Thomas Sampson, John Van Reenen	*The Costs and Benefits of Leaving the EU*
471	Dimitris Christelis, Dimitris Georgarakos, Tullio Jappelli	*Wealth Shocks, Unemployment Shocks and Consumption in the Wake of the Great Recession*
470	Gianmarco I. P. Ottaviano	*European Integration and the Gains from Trade*
469	Gianmarco I. P. Ottaviano, Filipe Lage de Sousa	*Relaxing Credit Constraints in Emerging Economies: The impact of public loans on the performance of Brazilian manufacturers*