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
Directorate General Research
Monetary Policy Research Division
European Central Bank

April 6, 2016

The paper has been published in the *Journal of Applied Econometrics*

DOI: 10.1002/jae.2514

Disclaimer

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.

Outline

Introduction

- The predictive likelihood is useful for **ranking models in forecast comparison exercises** using Bayesian inference.
- Geweke and Amissano (2010, p. 217) points out that the predictive likelihood function

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

- **Forecast problem:** an investigator is interested in comparing the performance wrt some but not all of the observable variables that can be predicted.
- **Remedy:** the observables which are **not** regarded as interesting must be integrated out from the predictive likelihood of the models where they appear.

Introduction

- This **marginalization problem** may be solved via textbook results when the joint predictive likelihood has a **known distributional form**.
- Such models are rare in practise. In the typical case when this distribution is unknown, we can 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 of the parameters) **times**
    - the **posterior density** wrt the model parameters.
- If the conditional likelihood is based on a distribution where marginalization can be handled analytically, the marginalization problem can be solved at this stage.
- **Gaussian:** Andersson and Karlsson (2008); Karlsson (2018); Geweke and Amisano (2010;2011 JoEconometrics;2012 AER); Amisano and Geweke (2013) have all computed the predictive likelihood via a Gaussian conditional likelihood.

- We suggest a **recursive approach**, based on the Kalman filter, to marginalize the conditional likelihood in linear Gaussian discrete-time state-space models. The approach builds up the marginalized parts of only the relevant arrays: **bottom-up** approach.
- It is therefore simpler than first calculating the mean and the covariance matrix of the joint conditional likelihood and thereafter reducing these arrays to the entries relevant for the marginalized conditional likelihood: **top-down** approach.
- The **methodological idea** of the paper can also be **extended** from the linear Gaussian world to **nonlinear and nonnormal models**, where a suitable particle filter can be applied.

- We apply the suggested approach to the forecast comparison exercise of Christoffel, Coenen and Warne (2011, CCW), where the focus is on the forecasting performance of the **New Area-Wide Model (NAWM) of the euro area** to reduced-form models.
- While the forecast comparison sample still begins in 1999Q1, the **endpoint is moved forward** from 2006Q4 to 2011Q4. Allows us to study how the models compare in terms of forecasting performance also during and after the onset of the **Great Recession**.
- We assess the results from a normal approximation of the predictive likelihood (as in CCW and Adolfson et al, 2007) to those obtained from an estimator based on **Monte Carlo integration**.
- We also include a **DSGE-VAR** model, with the NAWM as prior, in this setting, as well as to a **BVAR** (from CCW), and a vector **RW** model, all estimated with Bayesian methods for the same observables.

Outline

The Joint Predictive Likelihood

- For a sequence of future values $\mathcal{Y}_{T,h} = \{y_{T+1}, \dots, y_{T+h}\}$, where y_t is n-dimensional, the **joint predictive density of model** m

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

  $\mathcal{Y}_T^o$, in a recursive forecast comparison exercise. The **joint predictive likelihood of model** m is equal to the predictive density above **evaluated at the observed values** $\mathcal{Y}_{T,h}^o = \{y_{T+1}^o, \dots, y_{T+h}^o\}$.
- **Notice:** We may let y_t depend on m, but to simplify the notation we abstract from this here.

The Marginalized Predictive Likelihood

- Suppose we are interested in forecasting a **subset of the observable variables**, denoted by $\mathcal{Y}_{s,T,h} = \{y_{s,T+1}, \ldots, y_{s,T+h}\}$
- The marginalized predictive density is

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

- The **marginalized predictive likelihood** is given by this density evaluated at the observed values $\mathcal{Y}_{s,T,h}^o$.
- 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**.

Computing The Marginalized Conditional Likelihood

- In a linear Gaussian state-space model, the joint conditional likelihood is determined by its mean vector and covariance matrix. The same applies to the marginalized conditional likelihood.
- The predictive **mean vector** and the **covariance matrix** of the marginalized conditional likelihood can be **computed recursively via a Kalman filter** which allows for missing data. These "missing data" are identical to the variables we want to integrate out. Such a procedure gives a **bottom-up** approach to computing the marginalized conditional likelihood for each θ_m.
- Alternatively, the **top-down** approach is to compute the mean vector and the covariance matrix of the **joint conditional likelihood** and **remove** the elements of these objects that represent the variables we want to integrate out.

Estimating The Marginalized Predictive Likelihood

- Suppose we have N draws, $\theta_m^{(j)}$, from the posterior $p(\theta|\mathcal{Y}_T^o, m)$.
- A simple estimator of the marginalized predictive likelihood is obtained via **Monte Carlo (MC) integration**

$$\hat{p}_{MC}\big(\mathcal{Y}_{s,T,h}^o\big|\mathcal{Y}_T^o, m\big) = \frac{1}{N}\sum_{j=1}^{N} p\big(\mathcal{Y}_{s,T,h}^o\big|\mathcal{Y}_T^o, \theta_m^{(j)}, m\big)$$

- In practise, the MC estimator is expected to work well when the posterior draws cover well enough the parameter region where the marginalized conditional likelihood is large.
- Likelier when dimension of $\mathcal{Y}_{s,T,h}$ is fairly small and h is not too large.
- **Standard methods** for estimating the **marginal likelihood** may also be used.

Outline

Models in Empirical Application

- Like in CCW we compare forecasts using the **NAWM** of the ECB, a quite large log-linearized DSGE model of the euro area covering 18 observed variables.
- We also include a **DSGE-VAR** model for the same 18 variables, with the **NAWM as prior**, with $\lambda = 2.5$ and $p = 2$ lags. Does loosening the strong cross-equation restrictions of the NAWM improve the density forecasts?
- A **BVAR** model for the same variables, originally estimated by Marta Bańbura, which is built on the large BVAR methodology in Bańbura, Giannone and Reichlin (2010).
- A **vector random walk** model for the same variables, with a diffuse prior on the covariance matrix of the innovations. This model has a known distribution for the predictive density: **multivariate Student-**t.

Density Forecast Comparison Metric

- **Scoring rules** are used to **compare the quality of probabilitic forecasts** by giving a numerical value using the predictive distribution and an event of value that materializes.
- A widely used scoring rule is the **log predictive score**, suggested by, for example, Good (1952).

$$\mathcal{S}_{T+N_h+h-1}(m) = \sum_{t=T}^{T+N_h-1} \log p\left(\mathcal{Y}^o_{s,t,h} \middle| \mathcal{Y}^o_t, m\right),$$

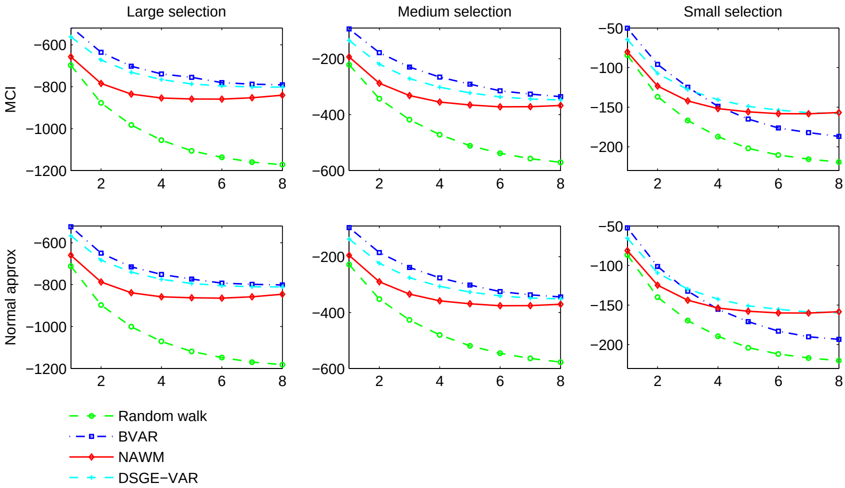
where $h = 1, \ldots, h^*$, $N_h = 1, \ldots, T_h$ is the number of time periods the h-step ahead predictive likelihood is evaluated.
- The **recursive average log predictive score** is

$$\bar{\mathcal{S}}_{T+N_h+h-1}(m) = \mathcal{S}_{T+N_h+h-1}(m)/N_h.$$

Sample and Subset of Variables

- The pseudo out-of-sample forecast ssample is 1999Q1–2011Q4. We only consider forecasts of quarterly growth rates for the variables that appear in first differences.
- We **exclude the five foreign variables** of the NAWM and **government consumption** from the comparisons since they are essentially exogenous in that model.
- Marginalized for each forecast horizon, we have three nested subsets:
  - **Small selection:** real GDP, GDP deflator inflation, interest rate.
  - **Medium selection:** Small selection and real private consumption, total investment, employment, nominal wages.
  - **Large selection:** Medium selection and exports, imports, import price deflator, private consumption deflator, real effective exchange rate.

MC Estimator of Log Predictive Score - Full Sample



Log predictive scores using the MC estimator of the log predictive likelihood and the normal approximation for the full sample 1999Q1–2011Q4

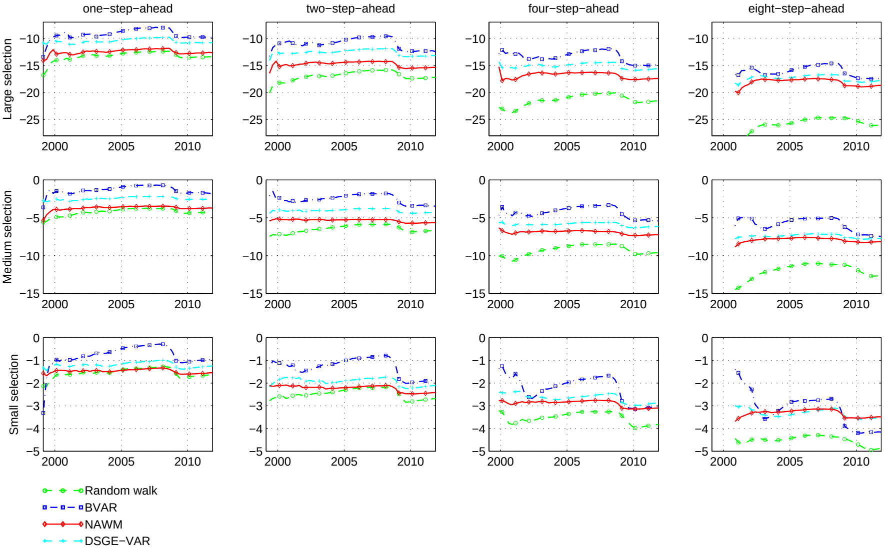
MC Estimator of Log Predictive Score - Full Sample

- **DSGE-VAR** generally obtains **higher log scores** than the **NAWM** for all horizons and variable selections. At the longer horizons, the NAWM comes nearer in performance to the DSGE-VAR.
- Taking into account model misspecification of the NAWM through a DSGE-VAR seems to improve the density forecasts.
- Compared with the **BVAR**, the **NAWM is outperformed** for the **large and medium selections** and all forecast horizons.
- For the **small selection**, the **forecast performance of the BVAR deteriorates** relative to the NAWM and the DSGE-VAR as the horizon increases, and the BVAR performs worse than these models for h-step-ahead forecasts over a year.
- The **normal approximation** gives nearly identical results as the MC estimator.

MC Estimator of Log Predictive Score - Full Sample

- CCW identify **two main factors** that may explain the relative strengths and weaknesses of the NAWM:
- Its **explicit microfoundations** give rise to a **parsimonious parametrized structure** with a large number of cross-equation restrictions. Potentially an advantage for achieving **forecast accuracy**.
- 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. 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 and GDP deflators. This also leads to a **systematic overprediction** of real private consumption growth.

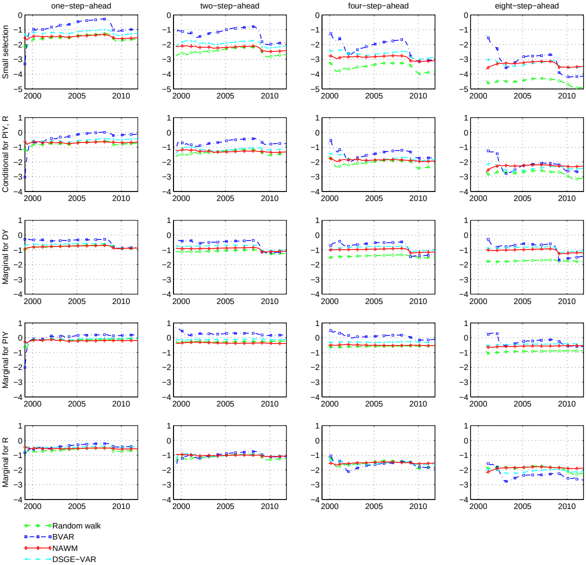
Recursive Average Log Score with MC Estimator



Recursive Average Log Score with MC Estimator

- For **large and medium selections**, the ranking of models over the various horizons is not greatly influenced by the choice of sample endpoint.
- The **average log scores** for 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 selections.
- The **performance of the BVAR deteriorates** substantially with the **onset of the Great Recession**, especially in the case of the **small selection**.
- The BVAR even looses its first rank position for the longer horizons. What may be the reason for this?

Recursive Average Log Score - Small Selection



BVAR: Recursive Posterior Mean of Real GDP Eq. Constant

Conclusions

- The paper discusses how the predictive likelihood can be computed, by means of marginalization, for any subset of the observed variables in linear Gaussian state-space models estimated with Bayesian methods.
- The suggest **bottom-up approach** based on the **Kalman filter** and combined with **MC integration** is applied in an extension of the CCW study for euro area data.
- **Compares the density forecasts** of the **NAWM**, a **DSGE-VAR** with the NAWM as prior, a **BVAR** based on the large BVAR methodology of Bańbura, Giannone and Reichlin (2010), and a **vector RW** model.
- **Model ranking:** the log predictive score (sum of the log predictive likelihood) **typically favors the BVAR** model, with the **DSGE-VAR improving somewhat on the density forecasts of the NAWM**, especially at the shorter horizons.

Conclusions

- For the **longer-term forecasts** and the **small selection** of variables (real GDP, GDP deflator, interest rate), the BVAR not only looses its first rank position to the DSGE-VAR at the onset of the Great Recession in 2008Q4, but also the second rank to the NAWM.
- The **main reason** appears to be the deterioration in the BVAR density forecasts of **real GDP growth** compared with those of the DSGE-VAR and the NAWM.
- 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-from BVAR model.
- The MC integration-based estimator is also compared with a normal approximation of the predictive likelihood. We find that **an assumption of a normal predictive density** provides a **good approximation of the predictive likelihood** of all four models.

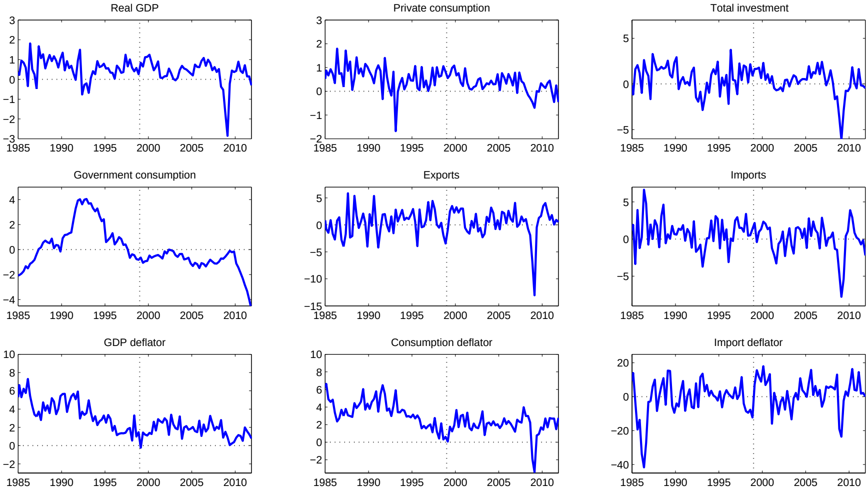
Outline

NAWM: The Data

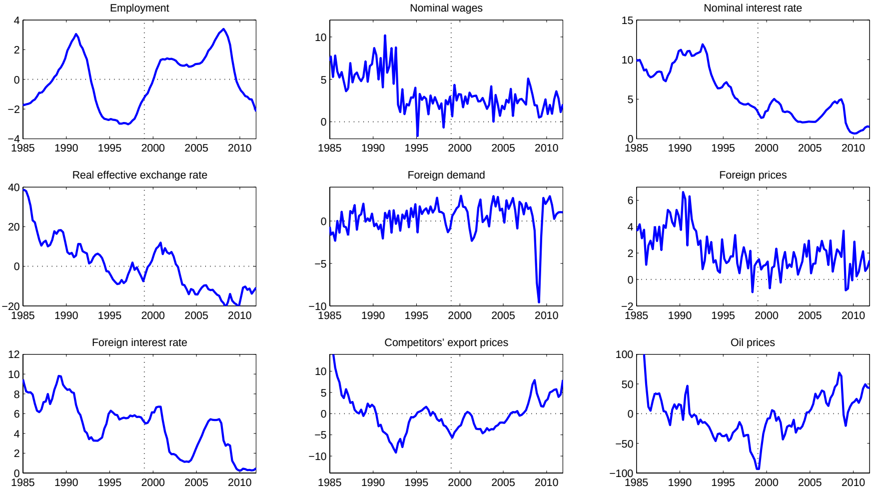
- We utilise data on 18 key macroeconomic times series:
  - Real GDP
  - Private consumption
  - Total investment
  - Extra-euro area exports
  - Extra-euro area imports
  - GDP deflator
  - Consumption deflator
  - Government consumption
  - Import deflator
  - Employment
  - Nominal wages
  - Nominal interest rate
  - Nominal effective exchange rate
  - Competitors' export prices†
  - Foreign demand†
  - Foreign GDP deflator†
  - Foreign nominal interest rate†
  - Oil price†

  covering the period 1985Q1-2011Q4 for the euro area (using 1980Q2-1984Q4 as training sample).
- The five times series marked with a dagger ('$\dagger$') are modelled using a structural VAR, while government consumption is modelled as an AR(2).

The Data: Part 1



The Data: Part 2



Some Estimation Details

- For the **NAWM** and the **DSGE-VAR** models we use **10,000 posterior draws** among the 500,000 post burn-in draws from the RWM sampler for each model and time period. These draws have been selected as draw number 1, 51, . . ., 499,951 to **combine modest computational costs** with a **lower correlation between draws**. Yields estimates of the log predictive likelihood that are accurate up to an including the first decimal. The posterior samples have been updated annually for these model.
- **Direct sampling** is possible for the **BVAR** through its normal-inverted Wishart posterior: we have used 50,000 draws when computing the predictive likelihood.
- For the **vector RW** model we have computed the predictive likelihood from its known distribution.