Coenen, Günter; Christoffel, Kai; Warne, Anders

Conference Paper

Forecasting with DSGE Models

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Forecasting Methods, No. A11-V1

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Coenen, Günter; Christoffel, Kai; Warne, Anders (2010) : Forecasting with DSGE Models, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Forecasting Methods, No. A11-V1, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:
https://hdl.handle.net/10419/37455

Forecasting with DSGE Models

Kai Christoffel, Günter Coenen and Anders Warne
DG-Research, European Central Bank

February 11, 2010

Abstract: In this paper we review the methodology of forecasting with log-linearised DSGE models using Bayesian methods. We focus on the estimation of their predictive distributions, with special attention being paid to the mean and the covariance matrix of h-steps ahead forecasts. In the empirical analysis, we examine the forecasting performance of the New Area-Wide Model (NAWM) that has been designed for use in the macroeconomic projections at the European Central Bank. The forecast sample covers the period following the introduction of the euro and the out-of-sample performance of the NAWM is compared to nonstructural benchmarks, such as Bayesian vector autoregressions (BVARs). Overall, the empirical evidence indicates that the NAWM compares quite well with the reduced-form models and the results are therefore in line with previous studies. Yet there is scope for improving the NAWM's forecasting performance. For example, the model is not able to explain the moderation in wage growth over the forecast evaluation period and, therefore, it tends to overestimate nominal wages. As a consequence, both the multivariate point and density forecasts using the log determinant and the log predictive score, respectively, suggest that a large BVAR can outperform the NAWM.

Keywords: Bayesian inference, DSGE models, euro area, forecasting, open-economy macroeconomics, vector autoregression.

JEL Classification Numbers: C11, C32, E32, E37.

Structural econometric forecasting, because it is based on explicit theory, rises and falls with the theory, typically with a lag.
Francis X. Diebold (1998, p. 175).

1. Introduction

Since the turn of the century, we have witnessed the development of a new generation of dynamic stochastic general equilibrium (DSGE) models that build on explicit micro-foundations with optimising agents. Major advances in estimation methodology allowed estimating variants of these models that are able to compete, in terms of data coherence, with more standard time series models, such as vector autoregressions (VARs); see, among others, Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2003, 2007), Adolfson, Laséen, Lindé, and Villani (2007), and Christoffel, Coenen, and Warne (2008). Accordingly, the new generation of DSGE

Note: The paper is in preparation for appearing as a chapter in an *'Oxford Handbook'* on Economic Forecasting, edited by Michael P. Clements and David F. Hendry. We are very grateful to Marta Bańbura who has estimated and computed the forecasts for the two large Bayesian VAR models we have used in the paper. We have received valuable comments from participants at the Nottingham workshop on DSGE modelling, December 2009, and seminar participants at the Humboldt University in Berlin. We are particularly grateful for discussions with and comments from Richard Anderson, Michael Burda and Alexander Meyer-Gohde. 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. Address to all authors: Directorate General Research, European Central Bank, Kaiserstrasse 29, 60311 Frankfurt am Main, Germany.

models provides a framework that appears particularly suited for evaluating the consequences of alternative macroeconomic policies.

Efforts have also been undertaken to bring these models to the forecasting arena. Results in Smets and Wouters (2004) suggest that the new generation of closed-economy DSGE models compare well with conventional forecasting tools such as VAR models; see also Edge, Kiley, and Laforte (2009). Similarly, the study by Adolfson, Lindé, and Villani (2007) shows that also open-economy DSGE models can compete well with reduced-form models. However, it is worth recalling that the study by Del Negro, Schorfheide, Smets, and Wouters (2007) finds evidence that the Smets and Wouters model is misspecified when estimated on postwar U.S. data and when applying the goodness-of-fit tools proposed in that study. Moreover, they show that this DSGE model is outperformed by a so-called DSGE-VAR in terms of out-of-sample point forecast accuracy.

Against this background, the goal of the current paper is to review and illustrate the methodology of forecasting with DSGE models using Bayesian methods. We limit the scope of the paper to *log-linearised* DSGE models; and, hence, we neither consider DSGE-VARs, as in Del Negro and Schorfheide (2004), nor do we consider DSGE models based on higher-order approximations, as in Fernández-Villaverde and Rubio-Ramírez (2005). As regards the initial steps of forecasting with DSGE models, Sargent (1989) was amongst the first to point out that a log-linearised DSGE model can be cast in the familiar state-space form, where the observed variables are linked to the model variables (and possibly to measurement errors) through the measurement equation. At the same time, the state equation provides the reduced form of the DSGE model, mapping current model variables to their lags and the underlying i.i.d. shocks, where the reduced form is obtained by solving for the expectation terms in the structural form of the model using a suitable method; see, e.g., Blanchard and Kahn (1980), Anderson and Moore (1985), Klein (2000), or Sims (2002). The Kalman filter can thereafter be used to compute the value of the log-likelihood function for any value of the model parameters when a (unique) solution of the DSGE model exists. A classical approach to the estimation of these parameters would then be to maximise the log-likelihood function with numerical methods. A Bayesian approach would instead complement the likelihood with a prior distribution for the parameters and estimate the posterior mode through numerical optimisation, or other properties of the posterior distribution via Markov Chain Monte Carlo (MCMC) methods.

In this paper, we shall discuss an algorithm for estimating the predictive distribution of the observed variables based on draws from the posterior distribution of the DSGE model parameters and simulation of future paths for the variables with the model. The general method, called sampling the future, was first suggested for univariate time series models by Thompson and Miller (1986). Their variant was simplified and adapted to VAR models by Villani (2001). The particular version of the algorithm that can be used for state-space models was suggested in Adolfson, Lindé, and Villani (2007). In case the forecast evaluation exercise only requires

moments from the predictive distribution, such as the mean and the covariance, then the simulation algorithm is not necessary. Estimation of such moments can instead be achieved by properly combining population moments for fixed parameter values with draws from the posterior distribution and, thus, without sampling the future via the model. However, if we also wish to estimate, e.g., quantiles, confidence intervals or the probability that the variables reach some barrier, then the simulation algorithm may prove useful. We note that the algorithm does not rely on a particular posterior sampler. It only requires that the draws characterise the posterior distribution of the parameters.

We illustrate these tools by applying them to a particular DSGE model. We have selected the New Area-Wide Model (NAWM), developed at the European Central Bank (ECB), which is designed for use in the (Broad) Macroeconomic Projection Exercises regularly undertaken by ECB/Eurosystem staff and for policy analysis. The specification of the NAWM was influenced by both economic and statistical criteria. For example, impulse-response functions and forecast-error-variance decompositions were used for assessing alternative specifications from an economic perspective, while the marginal likelihood and comparisons between model-based sample moments and estimates from the data only were applied as statistical model evaluation criteria. In addition, a small forecast evaluation exercise was conducted, but it was treated as one among many criteria for assessing the performance of the model. Here we extend the forecast evaluation exercise to the full set of the NAWM's endogenous variables. The forecast sample covers the period following the introduction of the euro and we shall study both point and density forecasts from 1 up to 8 quarters ahead. The DSGE model forecasts are compared to those from a VAR and three Bayesian VARs (BVARs), as well as the naïve random walk and (sample) mean benchmarks. We shall also consider different subsets of the observed variables included in the NAWM, as well as different transformations of these variables.

The remainder of the paper is organised as follows. Section 2 sketches the NAWM, while Section 3 reports on our implementation of Bayesian inference methods and on some selected estimation results for the NAWM. Section 4 first discusses how the predictive distribution of a DSGE model can be estimated, and it then presents the alternative forecasting models that are used in the empirical analysis. Section 5 covers the forecast evaluation of the NAWM, focusing first on point forecasts and then on density forecasts. Section 6 summarises the main findings of the paper and concludes.

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

In this section we provide a brief overview of the NAWM to set the stage for our review of the methodology for forecasting with log-linearised DSGE models. The NAWM is a micro-founded open-economy model of the euro area designed for use in the ECB/Eurosystem staff projections and for policy analysis; see Christoffel, Coenen, and Warne (2008) for a detailed description of the NAWM's structure. Its development has been guided by a principal consideration, namely to provide a comprehensive set of core projection variables, including a number of foreign variables,

which, in the form of exogenous assumptions, play an important role in the projections. As a consequence, the scale of the NAWM—compared with a typical DSGE model—is rather large, and it is estimated on 18 macroeconomic time series.

2.1. A Bird's Eye View on the Model

The NAWM features four classes of economic agents: households, firms, a fiscal authority and a monetary authority. Households make optimal choices regarding their purchases of consumption and investment goods, they supply differentiated labour services in monopolistically competitive markets, they set wages as a mark-up over the marginal rate of substitution between consumption and leisure, and they trade in domestic and foreign bonds.

As regards firms, the NAWM distinguishes between domestic producers of tradable differentiated intermediate goods and domestic producers of three types of non-tradable final goods: a private consumption good, a private investment good, and a public consumption good. The intermediate-good firms use labour and capital as inputs to produce their differentiated goods, which are sold in monopolistically competitive markets domestically and abroad. Accordingly, they set different prices for domestic and foreign markets as a mark-up over their marginal costs. The final-good firms combine domestic and foreign intermediate goods in different proportions, acting as price takers in fully competitive markets. The foreign intermediate goods are imported from producers abroad, who set their prices in euro, allowing for an incomplete exchange-rate pass-through. A foreign retail firm in turn combines the exported domestic intermediate goods, where aggregate export demand depends on total foreign demand.

Both households and firms face nominal and real frictions, which have been identified as important in generating empirically plausible dynamics. Real frictions are introduced via external habit formation in consumption and through generalised adjustment costs in investment, imports and exports. Nominal frictions arise from staggered price and wage-setting à la Calvo (1983), along with (partial) dynamic indexation of price and wage contracts. In addition, there exist financial frictions in the form of domestic and external risk premia.

The fiscal authority purchases the public consumption good, issues domestic bonds, and levies different types of distortionary taxes. Nevertheless, Ricardian equivalence holds because of the simplifying assumption that the fiscal authority's budget is balanced each period by means of lump-sum taxes. The monetary authority sets the short-term nominal interest rate according to a Taylor-type interest-rate rule, with the objective of stabilising inflation in line with the ECB's definition of price stability.

The NAWM is closed by a rest-of-the-world block, which is represented by a structural VAR (SVAR) model determining a small set of foreign variables: foreign demand, foreign prices, the foreign interest rate, foreign competitors' export prices and the price of oil. The SVAR model does not feature spill-overs from the euro area, in line with the treatment of the foreign variables as exogenous assumptions in the projections.

2.2. Some Key Model Equations

To better understand the cross-equation restrictions implied by the NAWM's structure, it is instructive to look at some key behavioural equations in their log-linearised form. We focus on those equations most closely related to the set of 12 observed variables that form the basis of the forecasting performance evaluation in Section 5; namely, private consumption, investment, imports and exports, the private consumption and the import deflator, wages and employment, the short-term nominal interest rate and the real effective exchange rate. Real GDP and the GDP deflator are obtained from the model's aggregate resource constraint in real and in nominal terms, respectively.

In order to derive the log-linearised equations, the NAWM is first cast into stationary form. To this end, all real variables are measured in per-capita terms and scaled by trend labour productivity z_t. This variable is assumed to follow a random walk with stochastic drift and defines the model's balanced growth path. Similarly, we normalise all nominal variables with the price of the consumption good $P_{C,t}$. For example, we use $c_t = C_t/z_t$ to denote the stationary level of per-capita consumption, while we use $p_{I,t} = P_{I,t}/P_{C,t}$ to represent the stationary relative price of the investment good. We then proceed with the log-linearisation of the transformed NAWM around its deterministic steady state, where the logarithmic deviation of a variable from its steady-state value is denoted by a hat ('$\widehat{\ }$'). For example, the log-deviation from steady state for the scaled consumption variable is $\widehat{c}_t = \log(c_t/c)$.

With these conventions, private consumption $\widehat{c}_t$ is characterised by an intertemporal optimality condition (Euler equation), which relates the log-difference of current and expected future consumption to the ex-ante real interest rate, $\widehat{r}_t - \mathrm{E}_t[\widehat{\pi}_{C,t+1}]$, noting that the specific form of the households' utility function, with additive habits and habit formation parameter κ, implies that also lagged consumption enters the consumption equation:

$$\begin{aligned}\widehat{c}_t \;=\; & \frac{1}{1+\kappa g_z^{-1}}\,\mathrm{E}_t\left[\widehat{c}_{t+1}\right] + \frac{\kappa g_z^{-1}}{1+\kappa g_z^{-1}}\,\widehat{c}_{t-1} - \frac{1-\kappa g_z^{-1}}{1+\kappa g_z^{-1}}\left(\widehat{r}_t - \mathrm{E}_t\left[\widehat{\pi}_{C,t+1}\right] + \widehat{\epsilon}_t^{RP}\right) \\ & - \frac{1}{1+\kappa g_z^{-1}}\left(\mathrm{E}_t\left[\widehat{g}_{z,t+1}\right] - \kappa g_z^{-1}\widehat{g}_{z,t}\right).\end{aligned} \tag{1}$$

Here, $\widehat{\epsilon}_t^{RP}$ denotes a risk-premium shock, which drives an exogenous wedge between the riskless interest rate set by the monetary authority and the effective interest rate faced by households. The expected quasi-difference of trend labour productivity growth, $\mathrm{E}_t[\widehat{g}_{z,t+1}] - \kappa g_z^{-1}\widehat{g}_{z,t}$, enters as an additional term because of the scaling of the consumption variable with the level of trend productivity, where g_z denotes the steady-state value of $g_{z,t} = z_t/z_{t-1}$.

Investment $\widehat{i}_t$ is characterised by an equation with a similar structure. The intertemporal price of investment is given by the log-difference of Tobin's Q—the discounted sum of expected future returns of the existing capital stock, with discount factor β—and the price of newly installed

capital goods, $\widehat{Q}_t - \widehat{p}_{I,t}$:

$$\widehat{i}_t = \frac{\beta}{1+\beta}\,\mathrm{E}_t\left[\widehat{i}_{t+1}\right] + \frac{1}{1+\beta}\,\widehat{i}_{t-1} + \frac{1}{\gamma_I\, g_z^2\,(1+\beta)}\left(\widehat{Q}_t - \widehat{p}_{I,t} + \widehat{\epsilon}_t^I\right) + \frac{1}{1+\beta}\left(\beta\,\mathrm{E}_t\left[\widehat{g}_{z,t+1}\right] - \widehat{g}_{z,t}\right). \tag{2}$$

The intertemporal price of investment is shifted by an investment-specific technology shock $\widehat{\epsilon}_t^I$, which affects the efficiency of newly installed capital goods. The lagged investment term reflects the existence of adjustment costs related to incremental changes in investment, with sensitivity parameter γ_I.

Private consumption and investment are composed of bundles of domestic and imported intermediate goods, $\widehat{im}_t^C$ and $\widehat{im}_t^I$. The demand for these import bundles depends on the total demand for the consumption good, $\widehat{q}_t^C = \widehat{c}_t$, and the investment good, $\widehat{q}_t^I = \widehat{i}_t$, respectively. Suppressing the consumption and investment superscripts for the sake of simplicity and focusing on the generic form of the import demand equation, the share of imports in total demand is then obtained as a function of the price of the imported intermediate-goods bundle relative to the price of the generic final good, $\widehat{p}_{IM,t} - \widehat{p}_t$:

$$\widehat{im}_t = -\mu\left(\widehat{p}_{IM,t} - \widehat{p}_t - \widehat{\Gamma}^{\dagger}_{IM,t}\right) + \widehat{q}_t. \tag{3}$$

Here, the parameter μ represents the price elasticity of import demand. As in the case of investment, adjustment costs are incurred which, in their generic form $\widehat{\Gamma}^{\dagger}_{IM,t}$, dampen the influence of changes in the relative price of imports on import demand.

The demand for euro area exports $\widehat{x}_t$ is determined in a similar way as a share of euro area foreign demand $\widehat{y}_t^*$. This share varies with the price of euro area exports (translated into foreign currency with the real effective exchange rate $\widehat{s}_t$, denominated in terms of the GDP deflator $\widehat{p}_{Y,t}$) relative to the price of exports of the euro area's competitors, $\widehat{p}_{X,t} - \widehat{s}_t - \widehat{p}_{Y,t} - \widehat{p}^{\,c}_{X,t}$:

$$\widehat{x}_t = -\mu^*\left(\widehat{p}_{X,t} - \widehat{s}_t - \widehat{p}_{Y,t} - \widehat{p}^{\,c}_{X,t} - \widehat{\Gamma}^{\dagger}_{X,t}\right) + \widehat{y}_t^* + \widehat{\nu}_t^*, \tag{4}$$

where the parameter μ^* denotes the price elasticity of exports. The term $\widehat{\Gamma}^{\dagger}_{X,t}$ represents generic adjustment costs, and the term $\widehat{\nu}_t^*$ is an exogenous shock to foreign export preferences.

Consumer prices are determined as a combination of the aggregate prices of the domestically produced and the imported intermediate goods, $\widehat{p}_{H,t}$ and $\widehat{p}_{IM,t}$. The evolution of these prices is governed, in generic form, by forward-looking Phillips-curve equations according to which the rate of price inflation $\widehat{\pi}_t$ gradually adjusts in response to fluctuations in real marginal costs $\widehat{mc}_t$, subject to an exogenous price mark-up shock $\widehat{\varphi}_t$:

$$\widehat{\pi}_t = \frac{\beta}{1+\beta\chi}\,\mathrm{E}_t\left[\widehat{\pi}_{t+1}\right] + \frac{\chi}{1+\beta\chi}\,\widehat{\pi}_{t-1} + \frac{(1-\beta\xi)\,(1-\xi)}{\xi\,(1+\beta\chi)}\left(\widehat{mc}_t + \widehat{\varphi}_t\right). \tag{5}$$

This equation derives from the typical Calvo assumption that firms can only infrequently re-set their prices optimally, namely with probability $1-\xi$. Those firms which are not permitted to do so are allowed to index their prices to past inflation $\widehat{\pi}_{t-1}$ with indexation parameter χ.

Real wages and hours worked are the key labour-market variables in the NAWM. Real wages $\widehat{w}_t$ adjust gradually according to a forward-looking Phillips-curve equation which closes the gap between the after-tax real wage $\widehat{w}_t^{\tau}$ and the marginal rate of substitution $\widehat{mrs}_t$, subject to an exogenous wage mark-up shock $\widehat{\varphi}_t^W$:

$$\begin{aligned}\widehat{w}_t &= \frac{\beta}{1+\beta}\,\mathrm{E}_t\left[\widehat{w}_{t+1}\right]+\frac{1}{1+\beta}\,\widehat{w}_{t-1}+\frac{\beta}{1+\beta}\,\mathrm{E}_t\left[\widehat{\pi}_{C,t+1}\right] \\ &\quad -\frac{1+\beta\chi_W}{1+\beta}\,\widehat{\pi}_{C,t}+\frac{\chi_W}{1+\beta}\,\widehat{\pi}_{C,t-1}-\frac{\left(1-\beta\xi_W\right)\left(1-\xi_W\right)}{\left(1+\beta\right)\xi_W\left(1+\frac{\varphi^W}{\varphi^W-1}\,\zeta\right)}\,\left(\widehat{w}_t^{\tau}-\widehat{mrs}_t-\widehat{\varphi}_t^W\right).\end{aligned} \tag{6}$$

As in the case of the price Phillips curves, the parameters $1-\xi_W$ and χ_W denote, respectively, the Calvo adjustment probability for (nominal) wages and the degree of indexation to past consumer price inflation $\widehat{\pi}_{C,t-1}$. The parameter φ^W denotes the steady-state wage markup and ζ the Frisch elasticity of labour supply.

Since there exist no reliable data for hours worked in the euro area, we rely on employment data and relate the employment variable $\widehat{E}_t$ to the NAWM's unobserved hours-worked variable $\widehat{N}_t$ by an auxiliary equation following Smets and Wouters (2003),

$$\widehat{E}_t = \frac{\beta}{1+\beta}\,\mathrm{E}_t[\widehat{E}_{t+1}]+\frac{1}{1+\beta}\,\widehat{E}_{t-1}+\frac{\left(1-\beta\xi_E\right)\left(1-\xi_E\right)}{\left(1+\beta\right)\xi_E}\,\left(\widehat{N}_t-\widehat{E}_t\right). \tag{7}$$

Here, the parameter ξ_E determines the sensitivity of employment with respect to hours worked, similar to the role of the Calvo parameters in the price and wage Phillips curves.

The monetary authority sets the short-term nominal interest rate $\widehat{r}_t$ according to a simple Taylor-type interest-rate rule, where the parameter ϕ_R represents the degree of interest-rate smoothing and the parameters ϕ_Π, $\phi_{\Delta\Pi}$ and $\phi_{\Delta Y}$ determine the sensitivity of the interest-rate response to, respectively, consumer price inflation, the change in inflation and real GDP growth (relative to trend productivity growth):

$$\widehat{r}_t = \phi_R\,\widehat{r}_{t-1}+\left(1-\phi_R\right)\phi_\Pi\,\widehat{\pi}_{C,t-1}+\phi_{\Delta\Pi}\left(\widehat{\pi}_{C,t}-\widehat{\pi}_{C,t-1}\right)+\phi_{\Delta Y}\left(\widehat{y}_t-\widehat{y}_{t-1}\right)+\widehat{\eta}_t^R. \tag{8}$$

The term $\widehat{\eta}_t^R$ denotes a serially uncorrelated monetary policy shock.

Finally, the real effective exchange rate $\widehat{s}_t$ is determined by a risk-adjusted uncovered interest parity condition:

$$\widehat{s}_t = \mathrm{E}_t\left[\,\widehat{s}_{t+1}\right]-\widehat{r}_t+\widehat{r}_t^*-\widehat{\epsilon}_t^{RP}+\mathrm{E}_t\left[\,\widehat{\pi}_{Y,t+1}-\widehat{\pi}_{Y,t+1}^*\right]-\gamma_{B^*}\,\widehat{s}_{B^*,t+1}-\widehat{\epsilon}_t^{RP^*}, \tag{9}$$

where $\widehat{r}_t^*$ and $\widehat{\pi}_{Y,t+1}^*$ denote the foreign interest rate and foreign inflation, respectively. The last two terms represent an external risk premium. It is composed of an endogenous component related to the net holdings of foreign bonds, $\widehat{s}_{B^*,t+1}$ with sensitivity γ_{B^*}, and an exogenous shock $\widehat{\epsilon}_t^{RP^*}$.

The NAWM's log-linearised equations, including the equations presented above, can be cast in state-space form, where the state equation corresponds to the reduced-form solution of the model, which we obtain using the AIM algorithm developed in Anderson and Moore (1985). The observed variables are related to the model's state variables through an appropriate measurement equation.

3. Bayesian Estimation of DSGE Models

We adopt the empirical approach outlined in Smets and Wouters (2003) and An and Schorfheide (2007) and estimate the NAWM employing Bayesian inference methods. This involves obtaining the posterior distribution of the model's parameters based on its log-linear state-space representation using the Kalman filter. For the empirical analyses, we use YADA, a Matlab programme for Bayesian estimation and evaluation of DSGE models; see Warne (2009).

In the following we sketch the adopted approach and describe the data and the shock processes that we consider in its implementation. We then briefly report on the calibration of the model's steady state and present some selected estimation results.

3.1. Methodology

Employing Bayesian inference methods allows formalising the use of prior information obtained from earlier studies at both the micro and macro level in estimating the parameters of a possibly complex DSGE model. This seems particularly appealing in situations where the sample period of the data is relatively short, as is the case for the euro area. From a practical perspective, Bayesian inference may also help to alleviate the inherent numerical difficulties associated with solving the highly non-linear estimation problem.

Formally, let $p(\theta_m|m)$ denote the prior distribution of the vector $\theta_m \in \Theta_m$ with structural parameters for some model $m \in \mathcal{M}$, and let $p(\mathcal{Y}_T|\theta_m, m)$ denote the likelihood function for the observed data, $\mathcal{Y}_T = \{ y_1, \ldots, y_T \}$, conditional on parameter vector θ_m and model m. The joint posterior distribution of θ_m for model m is then obtained by combining the likelihood function for $\mathcal{Y}_T$ and the prior distribution of θ_m,

$$p(\theta_m|\mathcal{Y}_T, m) \propto p(\mathcal{Y}_T|\theta_m, m)\, p(\theta_m|m),$$

where $\propto$ denotes proportionality.

The posterior distribution is typically characterised by measures of location, such as the mode or the mean, measures of dispersion, such as the standard deviation, or selected quantiles. Following Schorfheide (2000), we adopt an MCMC sampling algorithm to determine the joint posterior distribution of the parameter vector θ_m. More specifically, we rely on the random-walk Metropolis algorithm with a Gaussian proposal density to obtain a large number of random draws from the posterior distribution of θ_m. The posterior mode and the inverse Hessian matrix are computed by a standard numerical optimisation routine, namely Christopher Sims' optimiser `csminwel`.

As discussed in Geweke (1999), Bayesian inference also provides a framework for comparing alternative and potentially misspecified models on the basis of their marginal likelihood. For a given model m the latter is obtained by integrating out the parameter vector θ_m,

$$p(\mathcal{Y}_T|m) = \int_{\theta_m \in \Theta_m} p(\mathcal{Y}_T|\theta_m, m)\, p(\theta_m|m)\, d\theta_m.$$

Thus, the marginal likelihood gives an indication of the overall likelihood of the observed data conditional on a model. To estimate the marginal likelihood one may use the modified harmonic mean estimator, suggested by Geweke (1999); see also Geweke (2005). An alternative estimator, suggested by Chib and Jeliazkov (2001), relies on rewriting Bayes theorem into the so-called marginal likelihood identity. The former estimator requires only draws from the posterior of θ_m, while the latter also requires draws of these parameters from the proposal density.

3.2. Data and Shock Processes

In estimating the NAWM, we use times series for 18 macroeconomic variables which feature prominently in the ECB/Eurosystem staff projections: real GDP, private consumption, total investment, government consumption, extra-euro area exports and imports, the GDP deflator, the consumption deflator, the extra-euro area import deflator, total employment, nominal wages per head, the short-term nominal interest rate, the nominal effective exchange rate, foreign demand, foreign prices, the foreign interest rate, competitors' export prices, and the price of oil. All time series are taken from an updated version of the AWM database (see Fagan, Henry, and Mestre, 2005), except for the time series of extra-euro area trade data the construction of which is detailed in Dieppe and Warmedinger (2007). The sample period ranges from 1985Q1 to 2006Q4 (using the period 1980Q2 to 1984Q4 as training sample). The last five variables are modelled using an SVAR, the estimated parameters of which are kept fixed throughout the estimation of the NAWM. Similarly, government consumption is specified by means of an autoregressive (AR) process with fixed estimated parameters. For details, see Christoffel, Coenen, and Warne (2008, Section 3.2).

Prior to estimation, we transform real GDP, private consumption, total investment, extra-euro area exports and imports, the associated deflators, nominal wages per head, as well as foreign demand and foreign prices into quarter-on-quarter growth rates, approximated by the first difference of their logarithm. Furthermore, a number of additional 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 by removing the sample growth rate differentials, reflecting the fact that trade volumes and foreign demand tend to grow at a significantly higher rate than real GDP. Second, for the logarithm of government consumption we remove a linear trend consistent with the NAWM's steady-state growth rate of 2.0 percent per annum which is assumed to have two components: labour productivity growth,

g_z, of roughly 1.2 percent and labour force growth of approximately 0.8 percent. The former is broadly in line with the average labour productivity growth over the sample period. Third, we take the logarithm of employment and remove a linear trend consistent with a steady-state labour force growth rate of 0.8 percent, noting that, in the absence of a reliable measure of hours worked, we use data on employment in the estimation. Fourth, we construct a measure of the real effective exchange rate from the nominal effective exchange rate, the domestic GDP deflator and foreign prices (defined as a weighted average of foreign GDP deflators) and then remove the 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. Figure 1 shows the time series of the transformed variables for the sample period 1985Q1 to 2006Q4.

To ensure that the 1-step ahead covariance matrix in the likelihood function for the observed variables is non-singular, the NAWM features 12 distinct structural shocks, several of which have been discussed in Section 2.2 above, plus the 6 shocks in the AR and SVAR models for government consumption and the foreign variables, respectively. All shocks are assumed to follow first-order autoregressive processes, except for the monetary policy shock and the shocks in the AR and SVAR models, which are assumed to be serially uncorrelated. We recall in this context that assuming an autoregressive process for trend labour productivity growth $g_{z,t}$—referred to as the NAWM's permanent technology shock—implies that all real variables, with the exception of hours worked and employment, share a common stochastic trend, in line with the model's balanced-growth property.

In addition, we account for measurement error in extra-euro area trade data (both volumes and prices) in view of the fact that they are prone to revisions. We also allow for small errors in the measurement of real GDP and the GDP deflator to alleviate discrepancies between the national accounts framework underlying the construction of official GDP data and the NAWM's aggregate resource constraint.

3.3. Empirical Results

An extensive discussion of the empirical implementation of the NAWM is beyond the scope of this paper, and the reader is thus referred to Christoffel, Coenen, and Warne (2008) for details. Here we report selectively on the calibration of the model's steady state and the posterior distribution of some key estimated parameters, which is deemed helpful for understanding the model's forecasting performance analysed in Section 5.

Regarding the NAWM's steady state, all real variables are assumed to evolve along a balanced-growth path with a trend growth rate of 2 percent per annum, which roughly matches average real GDP growth in our estimation sample; see Table 1. Since the steady-state growth rate for the labour force can be seen as a proxy for population growth, all quantities within the NAWM can be interpreted in per-capita terms once it has been accounted for. Consistent with the balanced-growth assumption, we then calibrate key steady-state ratios of the model by matching their

empirical counterparts over the sample period. For example, the expenditure shares of private consumption, total investment and government consumption are set to, respectively, 57.5, 21 and 21.5 percent of nominal GDP, while the export and import shares are set to 16 percent, ensuring balanced trade in steady state. On the nominal side the monetary authority's long-run (net) inflation objective is set equal to 1.9 percent at an annualised rate, consistent with the ECB's quantitative definition of price stability of inflation being below, but close to 2 percent. This implies that, within the NAWM, nominal wages grow with a steady-state rate of 3.1 percent, corresponding to the sum of trend labour productivity growth of 1.2 percent and the inflation objective of 1.9 percent.

As to the choice of prior distributions for the NAWM's estimated parameters we follow Smets and Wouters (2003) since their closed-economy model of the euro area is essentially nested within the NAWM. Our choice of prior distributions for the parameters concerning the NAWM's open-economy dimension is informed by the priors employed in Adolfson, Laséen, Lindé, and Villani (2007). Comparing the plots of the prior and posterior distributions we find that the observed data provide additional information for most parameters. A number of estimation results are noteworthy. First, the estimates of the parameters shaping the dynamics of domestic demand in response to the model's structural shocks—the degree of habit formation in consumption, κ, and the investment adjustment cost parameter, γ_I—are broadly in line with those reported by Smets and Wouters. Second, on the nominal side, we observe that the estimate of the Calvo parameter constraining the frequency of price-setting decisions of domestic firms selling in home markets, ξ_H, is rather high. Yet our posterior mode estimate of about 0.92 is comparable with a point estimate of about 0.90 for the Calvo parameter in the model of Smets and Wouters. The estimate implies that the NAWM's domestic Phillips curve is rather flat or, in other words, that the sensitivity of domestic inflation with respect to movements in real marginal cost is low. Similarly, the posterior mode estimate of the indexation parameter χ_H is 0.42, suggesting a relatively low degree of inflation persistence. Third, regarding the interest-rate rule, we observe that the estimated response coefficients ϕ_R, ϕ_Π, $\phi_{\Delta\Pi}$ and $\phi_{\Delta Y}$ are rather close to the estimates reported in Smets and Wouters, despite the fact that the NAWM's interest-rate rule does not feature a response to the so-called flex-price output gap, unlike the rule considered by Smets and Wouters. Finally, regarding the properties of the structural shocks, none of the estimated shock processes appears excessively persistent.

Figure 2 depicts the prior and posterior distributions of the structural parameters κ, γ_I, ξ_H and χ_H, and the response coefficients of the interest-rate rule, ϕ_R, ϕ_Π, $\phi_{\Delta\Pi}$ and $\phi_{\Delta Y}$, using the full sample, whereas Figure 3 shows the sequence of the posterior mode estimates when the sample is updated recursively over the period following the introduction of the euro. Overall, the recursively updated posterior mode estimates reveal a rather high degree of stability. Yet the gradual upward shift of the Calvo parameter ξ_H suggests that domestic inflation has become less sensitive to movements in marginal costs over time. The gradual fall in the indexation

parameter χ_H implies a diminishing degree of inflation persistence, which may be interpreted as an indication that the anchoring of inflation expectations has been strengthened with the introduction of the euro area.

4. Bayesian Forecasting by Sampling the Future

4.1. Estimating the Predictive Distribution of a DSGE Model

Let $\theta \in \Theta$ be a vector of parameters for the log-linearised DSGE model; to simplify notation we have omitted the model m index in this section. Given that a unique convergent solution exists at a particular value for the parameter vector, we can express the relationship between the model variables, defined as deviations from the steady state, and the parameters as a VAR system. Specifically, let η_t be 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 for $t = 1, 2, \ldots, T$. The solution (reduced form) of a log-linearised DSGE model can now be represented by:

$$\xi_t = F\xi_{t-1} + B\eta_t, \quad t = 1, \ldots, T, \tag{10}$$

where F and B are uniquely determined by θ. The observed variables are denoted by y_t, an n-dimensional vector, which is linked to the model variables ξ_t through the equation

$$y_t = A'x_t + H'\xi_t + w_t, \quad t = 1, \ldots, T. \tag{11}$$

The k-dimensional vector x_t is here assumed to be deterministic, while w_t is a vector of i.i.d. normal measurement errors with mean zero and covariance matrix R. The measurement errors and the shocks η_t are assumed to be independent, while the matrices A, H, and R are uniquely determined by θ.

The system in (10) and (11) is a state-space model with ξ_t being partially unobserved state variables when, for example, $r > n$. Equation (10) gives the state or transition equation and (11) the measurement or observation equation. Provided the number of measurement errors and structural shocks is large enough, we can calculate the likelihood function for the observed data $\mathcal{Y}_T = \{y_1, \ldots, y_T\}$ via the Kalman filter; see, e.g., Hamilton (1994) for details. The filter can also be used to estimate all unobserved variables in the model at the given value for θ.

The predictive density of $y_{T+1}, \ldots, y_{T+H}$ can be expressed as

$$p\big(y_{T+1}, \ldots, y_{T+H}|\mathcal{Y}_T\big) = \int_{\theta\in\Theta} p\big(y_{T+1}, \ldots, y_{T+H}|\mathcal{Y}_T, \theta\big)p\big(\theta|\mathcal{Y}_T\big)d\theta, \tag{12}$$

where $p(\theta|\mathcal{Y}_T)$ is the posterior density of θ based on the data available at time T. Since the integral in (12) cannot be evaluated analytically we can apply a numerical algorithm adapted by Adolfson, Lindé, and Villani (2007) to state-space models; see also Thompson and Miller (1986). That is:

(1) Draw θ from $p(\theta|\mathcal{Y}_T)$;

(2) Draw the state variables at time T from $\xi_T \sim N(\xi_{T|T}, P_{T|T})$, where $\xi_{T|T}$ is the filter estimate of ξ_T and $P_{T|T}$ is the covariance matrix of ξ_T given θ and $\mathcal{Y}_T$;

(3) Simulate a path for the state variables from (10) using the drawn value for ξ_T as initial value and a sequence of structural shocks $\eta_{T+1}, \ldots, \eta_{T+H}$ drawn from $N(0, I_q)$;

(4) Draw a sequence of measurement errors $w_{T+1}, \ldots, w_{T+H}$ from $N(0, R)$ and compute the path for the observed variables $y_{T+1}, \ldots, y_{T+H}$ using the measurement equation (11);

(5) Repeat steps 2-4 M_1 times for the same θ;

(6) Repeat steps 1-5 M_2 times.

The algorithm thus gives $M = M_1 M_2$ paths from the predictive distribution in (12). Point and interval forecasts as well as quantiles can now be computed in a straightforward manner. However, it may be noted that if the forecast evaluation exercise only requires moments from the predictive distribution, such as the mean and the covariance matrix, then the above algorithm is not needed. The population mean of y_{T+h} given $\mathcal{Y}_T$ and θ is

$$E\big[y_{T+h}|\mathcal{Y}_T, \theta\big] = A' x_{T+h} + H' F^h \xi_{T|T}, \quad h = 1, \ldots, H. \tag{13}$$

To estimate the mean of the predictive distribution of y_{T+h} we may simply compute the sample average of the right hand side of (13) for $\theta^{(i)} \sim p(\theta|\mathcal{Y}_T)$, $i = 1, \ldots, M$. By choosing M large enough, the numerical standard error of this estimator of $E[y_{T+h}|\mathcal{Y}_T]$ is negligible.

Similarly, the covariance matrix of y_{T+h} conditional on $\mathcal{Y}_T$ and θ is

$$C\big[y_{T+h}|\mathcal{Y}_T, \theta\big] = H' F^h P_{T|T} \big(F^h\big)' H + H' \left(\sum_{j=1}^{h} F^{j-1} B B' \big(F^{j-1}\big)' \right) H + R. \tag{14}$$

The first term on the right hand side represents state-variable uncertainty given θ, the second term reflects uncertainty due to the structural shocks, and the third the uncertainty due to measurement errors. Following Adolfson, Lindé, and Villani (2007), the prediction covariance matrix of y_{T+h} is given by

$$C\big[y_{T+h}|\mathcal{Y}_T\big] = E_T\left[C\big[y_{T+h}|\mathcal{Y}_T, \theta\big]\right] + C_T\left[E\big[y_{T+h}|\mathcal{Y}_T, \theta\big]\right], \tag{15}$$

where E_T and C_T denote the expectation and covariance with respect to the posterior of θ at time T. The second term on the right hand side of (15) measures the impact that parameter uncertainty has on the h-steps ahead forecasts based on the population mean, while the first term can be decomposed into uncertainties due to unobserved state variables, structural shocks and measurement errors, where the dependence on the parameters has now been dealt with. The first term in (15) can be estimated by the sample average of $C[y_{T+h}|\mathcal{Y}_T, \theta^{(i)}]$ in (14) for the M draws from $p(\theta|\mathcal{Y}_T)$, while the second term can be estimated by the sample covariance matrix of $E[y_{T+h}|\mathcal{Y}_T, \theta^{(i)}]$ in (13) using these M draws. Again, we can choose M large enough such that the numerical standard errors of the estimators are negligible.

4.2. Alternative Forecasting Models

Sims (1980) convincingly argued that VARs provide a less restrictive environment for modelling macroeconomic time series than the large-scale structural macroeconometric models, based on incredible identifying assumptions, that were prevalent at the time. However, while VARs often provide a reasonably good fit of macroeconomic time series data, a problem with using them is that they are not parsimonious and, hence, the number of variables that can be included is limited by a lack of long time series. To overcome this problem in forecasting situations, the so called Minnesota prior (Doan, Litterman, and Sims, 1984) makes use of the old idea of shrinkage, a flexible method for constraining the dimension of the parameter space. Given the view that the random walk is relatively accurate for forecasting macroeconomic time series (in levels), the Minnesota prior is based on shrinking the VAR parameters towards univariate random-walk processes.

Moreover, VAR models may be considered as linear approximations of DSGE models. For instance, using the idea that VARs can be used to summarise the statistical properties of both observed time series data and data simulated from a DSGE model, Smith (1993) showed how they can serve as a device from which the structural parameters could be estimated and for conducting (indirect) inference; see also Gouriéroux, Monfort, and Renault (1993). Furthermore, the state-space representation in (10)-(11) can, under certain conditions, be rewritten as an infinite order VAR model; see Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007). If these conditions are not met, then the state-space representation of the DSGE model may have a VARMA representation, where the moving average term is not invertible.[1]

An early attempt of combining DSGE models with Bayesian VARs is Ingram and Whiteman (1994), who proposed a way of deriving priors for VARs from the economic model. This approach was further developed by Del Negro and Schorfheide (2004) into the so-called DSGE-VAR, where the DSGE model is used to determine the moments of the prior distribution of the VAR parameters using a normal/inverted Wishart form. The authors found that this model can compete in forecasting exercises with BVARs based on the Minnesota prior. Similar to the ideas in Smith (1993), they demonstrated how posterior inference about the DSGE model parameters can be conducted via the VAR by integrating out the dependence of the VAR parameters from the joint posterior and thereby obtaining a marginal likelihood function for the parameters of the DSGE model; see also Del Negro and Schorfheide (2006). Moreover, they showed how the DSGE model can be utilised for providing identifying restrictions for the DSGE-VAR, thereby allowing for comparisons of, e.g., impulse responses between the DSGE model and the DSGE-VAR. The DSGE-VAR approach was further enriched by Del Negro, Schorfheide, Smets, and Wouters (2007) into a framework for assessing the time series fit of a DSGE model.

[1] Since the number of shocks and measurement errors of the NAWM is greater than the number of observed variables, the model does not satisfy the conditions in Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007).

In this study, we shall consider two classes of BVARs, one that is intended for systems with a smaller dimension and one that has been proposed for large data sets; cf. Bańbura, Giannone, and Reichlin (2008). The usefulness of BVARs of the Minnesota type for forecasting purposes has long been recognised, as documented early on by Litterman (1986), and such models are therefore natural benchmarks in forecast evaluations. While a DSGE-VAR is also a relevant candidate forecast model, we have opted to focus on BVARs with statistically motivated priors. In addition to models estimated with Bayesian methods, we shall also consider more traditional forecasting models in our empirical exercise.

The small BVAR is based on the parameterisation and prior studied by Villani (2009). That is, we consider a VAR model with a prior on the steady-state parameters, and a Minnesota-style prior on the parameters on the lags of the endogenous variables; see also Adolfson, Lindé, and Villani (2007). For the p-dimensional covariance stationary vector z_t the VAR is given by:

$$z_t = \Psi d_t + \sum_{l=1}^{k} \Pi_l \big(z_{t-l} - \Psi d_{t-l}\big) + \varepsilon_t, \quad t = 1, \ldots, T. \tag{16}$$

The d-dimensional vector d_t is deterministic, and the residuals ε_t are assumed to be i.i.d. normal with zero mean and positive definite covariance matrix Ω. The Π_l matrix is $p \times p$ for all lags, while Ψ is $p \times d$ and measures the expected value of x_t conditional on the parameters and other information available at $t = 0$.

One advantage with the parameterisation in (16) is, as pointed out by Villani (2009), that the steady state (or mean) of the endogenous variables is directly parameterised via Ψ. For the standard parameterisation of a VAR model the parameters on the deterministic variables are written as $\Phi = (I_p - \sum_{l=1}^{k} \Pi_l)\Psi$ when $d_t = 1$. This makes it difficult to specify a prior on Φ which gives rise to a reasonable prior distribution on the steady state. Moreover, when z_t is a subset of the observed variables used in the estimation of the NAWM, we can directly form a prior on the steady state of z_t that is consistent with the steady-state prior for the NAWM as captured by a prior on A. This allows for a more balanced comparison between the models since they can share the same prior mean, or steady state, for the variables that appear in both models. The steady state in the NAWM is calibrated, while the steady-state prior covariance matrix is positive definite for the BVAR. Hence, some imbalance between the models remains for the steady-state parameters. Details on the small BVAR model specification are given in Appendix A.

Let $p(\Psi, \Pi, \Omega | \mathcal{Z}_T)$ denote the posterior density, where $\Pi = [\Pi_1 \ \cdots \ \Pi_k]$ and $\mathcal{Z}_T = \{z_1, \ldots, z_T\}$. Simulation from this distribution is performed via Gibbs sampling for the three groups of parameters Ψ, Π, and Ω using the full conditional posteriors given by Villani (2009, Proposition A.1). Out-of-sample forecasts for the BVAR are calculated for the sample $T+1, \ldots, T+H$, with the objective of estimating the predictive distribution $p(z_{T+1}, \ldots, z_{T+H} | \mathcal{Z}_T)$. The algorithm used for a BVAR was adapted to a multivariate setting by Villani (2001) from the univariate approach suggested by Thompson and Miller (1986). That is,

(1) Draw (Ψ, Π, Ω) from $p(\Psi, \Pi, \Omega|\mathcal{Z}_T)$;

(2) Draw residuals $\varepsilon_{T+1}, \ldots, \varepsilon_{T+H}$ from $N(0, \Omega)$ and calculate a path for the endogenous variables $z_{T+1}, \ldots, z_{T+H}$ using the VAR in (16);

(3) Repeat step 2 M_1 times for the same (Ψ, Π, Ω);

(4) Repeat steps 1-3 M_2 times.

If the forecast evaluation exercise only requires estimates of, e.g., the mean and the covariance matrix of the predictive distribution, the above algorithms need not be used. For example, if the lag order of the VAR is $k = 1$ and $d_t = 1$, the mean of z_{T+h} given $\mathcal{Z}_T$ and the parameters is

$$E\big[z_{T+h}|\mathcal{Z}_T, \Psi, \Pi, \Omega\big] = \Psi + \Pi^h\big(z_T - \Psi\big). \tag{17}$$

The mean of z_{T+h} given $\mathcal{Z}_T$ can therefore be estimated by the average of the right-hand side of (17) over M draws from the posterior of (Ψ, Π, Ω). To estimate the covariance matrix of the predictive distribution we first note that

$$C\big[z_{T+h}|\mathcal{Z}_T, \Psi, \Pi, \Omega\big] = \sum_{i=0}^{h-1} \Pi^i \Omega \left(\Pi^i\right)'. \tag{18}$$

The covariance matrix of z_{T+h} given $\mathcal{Z}_T$ can now be estimated by adding the sample average of (18) over M draws from the posterior of the parameters to the sample covariance matrix of (17) over the same draws. The former term measures the part of the h-steps ahead forecast uncertainty due to the VAR innovations, while the latter term reflects parameter uncertainty.

In this paper, the variables in the BVAR with a steady-state prior are the same as those that were used by Smets and Wouters (2003), except for that they are measured as in the NAWM. That is, we use the following variables: real GDP growth, real private consumption growth, real total investment growth, GDP deflator inflation, employment, nominal wage growth, and the short-term nominal interest rate. Hence, two of the variables are given in levels (employment and the short-term nominal interest rate), while the remaining appear in first differences.

Bańbura, Giannone, and Reichlin (2008) advocate the use of high-dimensional BVARs for macroeconomic forecasting purposes. Building on the well-known Minnesota prior and its developments (Doan, Litterman, and Sims, 1984; Litterman, 1986), the authors suggest that as the dimension of the model increases, the overall shrinkage should be stronger; i.e., that the prior should be tighter. Building on this idea, the authors find that the forecasting performance of a small VAR model can be much improved upon by considering a high-dimensional VAR model (131 macroeconomic indicators). Moreover, their results suggest that forecasting performance is already substantially improved when the VAR model has 20 (carefully) selected macroeconomic variables.

We will therefore include two large BVARs that cover the same 18 variables as the NAWM in the study. That is, we let $d_t = 1$ and $z_t = y_t$ so that $p = n$ in (16). Moreover, we reparameterise the deterministic part such that we can use the constant term (Φ) instead of the steady-state

term (Ψ). The VAR may therefore be expressed as:

$$y_t = \Phi + \sum_{l=1}^{k} \Pi_l y_{t-l} + \varepsilon_t, \quad t = 1, \ldots, T. \tag{19}$$

The prior distribution is based on the extension of the usual Minnesota prior to a normal/inverted Wishart, as in Kadiyala and Karlsson (1997) and Robertson and Tallman (1999), and this prior is implemented via dummy observations (see, e.g., Lubik and Schorfheide, 2006). Additional dummy observations are added through a prior on the sum of the Π_l matrices, thereby yielding non-zero prior correlations between the autoregressive parameters (see Sims and Zha, 1998). Details concerning the implementation of the dummy observations prior are given in Bańbura, Giannone, and Reichlin (2008); see also Appendix B below.

Two large BVAR models for y_t will be studied in the forecast exercises below. These models primarily differ in how the prior mean of the autoregressive parameters is treated. In both models, the prior mean of Π_l for all $l \geq 2$ as well as for the off-diagonal elements of Π_1 are zero. For the diagonal elements of Π_1, the prior mean is zero in one of the large BVAR models, henceforth the white-noise prior. The second large BVAR sets the prior mean of these diagonal elements equal to unity if the variable is measured in levels, and zero if in first differences. Below we shall refer to this as a mixed prior. Apart from these differences in the treatment of the mean, the priors of the two large BVARs differ only in terms of the numeric value given to the overall tightness hyperparameter; cf. Appendix B.

Posterior sampling is straightforward for the large BVAR models. Specifically, the marginal posterior of Ω is inverted Wishart, while the posterior distribution of (Φ, Π) conditional on Ω is normal; see Appendix B for details. To sample from the joint posterior we may therefore use direct sampling; see, e.g., Geweke (2005, Chapter 4.1).

Since we will compare the forecasting performance of the NAWM with a small BVAR, we shall also estimate a VAR model for the same choice of variables in z_t with maximum likelihood, with the same lag length as all the BVARs ($k = 4$). Moreover, we shall check how well the DSGE model fares when comparing it to the naïve random-walk and mean benchmarks. The mean is here estimated by the within-sample mean of the variables to be forecast. Similarly, we shall consider a random walk in the variables that are forecasted. Below we shall study the forecasting performance for both quarterly and annual changes of (a subset of) the variables that appear in first differences in the NAWM. Hence, the NAWM and the various VAR models do not change with these changes in the forecasted variables (although their forecasts are affected by it), the mean and the random-walk models do change. Accordingly, no matter which criterium is used for evaluating the forecasting performance across annual and quarterly changes, the ranking of the mean and random-walk models relative to the other models is likely to change.[2]

[2] If a variable x_t appears in first differences in the NAWM, $\Delta x_t = x_t - x_{t-1}$, the random-walk model for quarterly changes is simply $\Delta x_t = \Delta x_{t-1} + \epsilon_t$, while the random-walk model for annual changes is $\Delta_4 x_t = \Delta_4 x_{t-1} + \epsilon_t$. The latter model can be rewritten as $\Delta x_t = \Delta x_{t-4} + \epsilon_t$. Similarly, the mean model for quarterly changes is

5. Evaluating Forecast Accuracy

The forecast performance of the NAWM along with the 6 reduced-form models will be assessed in this section using a rolling procedure where the parameters are estimated up to period T when the predictive distribution of periods $T+1,\ldots,T+H$ is to be computed and when T is the 4th quarter of the year. When T corresponds to some quarter $i=1,2,3$, the DSGE model and the alternative Bayesian models are estimated using data until $T-i$. Hence, these models are re-estimated annually. The other models are always estimated with data until period T.

The first out-of-sample forecasts are computed for 1999Q1, i.e., the first quarter after the introduction of the euro, while the final period is 2006Q4. The length of the maximum forecast horizon, H, is 8 quarters, yielding 32 observations of the 1-step ahead forecasts and 25 of the 8-steps ahead forecasts. Most variables in the NAWM, such as real GDP, are measured in first differences at a quarterly frequency. Since year-on-year changes are often of interest in practice we shall also, as mentioned above, study how the models perform when forecasting annual changes.

The forecast comparisons involve both point forecasts and density forecasts. For the point forecasts we analyse univariate and multivariate mean squared error (MSE) measures. The univariate tool is the usual root mean square error, while the trace and log determinant statistics of scaled MSE matrices for the different horizons are used when examining multivariate point forecasts. For the density forecasts we focus on the log predictive score under the assumption of normality for each individual forecast horizon.

5.1. Point Forecasts

Figure 4 shows the root mean squared forecast errors (RMSE) when forecasting quarterly changes for the variables in first differences. To facilitate the comparisons with the multivariate point forecast analysis below, the forecast errors have here been scaled with the estimated standard deviation of the variable over the period 1995Q1-2006Q4.

The number of variables in Figure 4 is equal to 12 and the variables are the same as the ones we shall focus on both for the multivariate point forecasts and the density forecasts. The remaining 6 variables are essentially exogenous for the NAWM and they include the 5 variables in the foreign SVAR block and real government consumption. Since the parameters that determine the behaviour of these variables have been calibrated using data until 2006Q4, the comparison between the NAWM and the 6 alternative models in an out-of-sample forecasting exercise would not be fair in those dimensions and we have therefore excluded them from the analysis.

The univariate RMSE analysis reveals that the DSGE model fares quite well against the competitors. In particular, the NAWM does well in forecasting real GDP growth, real export and import growth, import price deflator inflation, employment and the short-term nominal interest

$\Delta x_t = \mu_q + \epsilon_t$, while the mean model for annual changes is $\Delta_4 x_t = \mu_a + \epsilon_t$. The latter model can equivalently be expressed as $\Delta x_t = \mu_a - \Delta x_{t-1} - \Delta x_{t-2} - \Delta x_{t-3} + \epsilon_t$.

rate. The most difficult dimensions for the DSGE model concern nominal wage growth in particular, but also consumption deflator inflation at the shorter horizons. It is worth underlining that all forecast models have dimensions where their performance is relatively good, and dimensions where they are less successful.

The RMSE results when forecasting annual changes of the variables in first differences are shown in Figure 5; the results for employment, the nominal interest rate and the real effective exchange rate concern their levels and are therefore equal to those in Figure 4. It should be noted that the scaling of the RMSEs are now based on the estimated standard deviations for *annual* real GDP growth, etc., and hence the scaled RMSEs for, e.g., the 1-step ahead forecasts differ from the results when forecasting quarterly changes. Moreover, recall that the mean and random-walk models change when forecasting annual rather than quarterly growth since they concern the mean of, e.g., annual real GDP growth and a random walk in annual real GDP growth.

It appears from Figure 5 that the naïve random-walk benchmark performs better relative to the competitors when forecasting annual changes instead of quarterly changes. In particular, it seems to work rather well when forecasting annual changes in private consumption, the GDP deflator, the consumption deflator, and nominal wages; see Atkeson and Ohanian (2001) for a discussion of the forecasting accuracy of annual inflation when using Phillips curves relative to a random walk on U.S. data. The more successful dimensions of the DSGE model remain when forecasting annual changes, while some of its weaker dimensions are emphasised. For instance, the NAWM does not forecast annual real private consumption growth well beyond 2 quarters when compared with, e.g., the random-walk model.

Multivariate measures of point forecast accuracy are often based on the (scaled) h-steps ahead MSE matrix:

$$\Sigma_M(h) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \tilde{\epsilon}_{t+h|t} \tilde{\epsilon}'_{t+h|t}, \quad h = 1, \ldots, H, \tag{20}$$

where $\tilde{\epsilon}_{t+h|t} = M^{-1/2} \epsilon_{t+h|t}$, and $\epsilon_{t+h|t}$ is the h-steps ahead forecast error from a forecast produced at t. The scaling matrix M is positive definite, while N_h is the number of h-steps ahead forecasts.

The trace and the log determinant are two measures that are often used in practice for evaluating multivariate forecast accuracy; see, e.g., Adolfson, Lindé, and Villani (2007). The choice of scaling matrix has a direct impact on the ranking of forecasting models when using the trace statistic since $\text{tr}[\Sigma_M(h)] = \text{tr}[M^{-1}\Sigma_I(h)]$, where $\Sigma_I(h)$ is based on $M = I$, the identity matrix. Since $\log|\Sigma_M(h)| = \log|\Sigma_I(h)| - \log|M|$ it follows that the log determinant statistic is invariant to the choice of M.

Moreover, and as emphasised by Clements and Hendry (1993), measures based on the MSE matrix in (20) are, at least, for linear models generally *not* invariant to non-singular, scale-preserving linear transformations, while the class of models is itself invariant to such isomorphic transformations. This is due to the fact that the h-steps ahead forecast errors are linear functions of current and past innovations up to order $h-1$. The chosen transformation affects the

parameterisation of these forecast errors and, thus, the weights given to the innovations. An exception is the log determinant statistic for the 1-step ahead forecasts, but not when $h \geq 2$. Since MSE-based measures are unable to account for the correlation between forecast errors at different horizons, the ranking of forecast models may depend on the choice of data transformation.[3] When we compute forecasts of, say, annual changes from a model with variables in quarterly changes, the resulting forecast errors for the annual changes are equal to the sum of the forecast errors for quarterly changes for the current and previous three quarters. Hence, MSE-based statistics may lead to a different ranking of models when forecasting annual changes compared with quarterly changes since the weights on the innovations in the moving average expressions of the forecast errors are affected by the choice of transformation.

The trace and the log determinant statistics are both functions of the eigenvalues of the MSE-matrix, where the largest eigenvalue gives the least predictable dimension and the smallest the most predictable. Since the trace is equal to the sum of the eigenvalues it follows that this statistic tends to be dominated by the largest eigenvalues, while the determinant is the product of the eigenvalues and is therefore also influenced by the smallest. That is, the trace measure tends to be dominated by the least predictable dimensions, while the log determinant measure may be driven by the most predictable dimensions.

The MSE statistics are computed for 3 different cases. First, we consider all the 12 variables displayed in Figure 4. Since the small VAR and BVAR models do not cover all these variables, the second case has the 7 variables that all models cover. That is, real GDP growth, real private consumption growth, real total investment growth, GDP deflator inflation, employment, nominal wage growth, and the short-term nominal interest rate. Finally, we examine a case with only 3 of these 7 variables, namely, real GDP growth, GDP deflator inflation and the short-term nominal interest rate. This choice of variables may be viewed as comprising the minimum set of variables relevant to monetary policy analysis.

The trace statistics when forecasting quarterly changes for the variables in first differences are displayed in Figure 6. The scaling matrix M is here assumed to be diagonal with diagonal elements given by the variances of the variables. The variances have been estimated over the period 1995Q1-2006Q4 and the scaling is therefore the same as for the individual RMSEs in Figure 4.

With this scaling matrix, the trace statistics are the sum of the squared RMSEs in Figure 4. Given the results for the univariate RMSEs when forecasting quarterly changes, it is therefore not surprising that the NAWM compares favourably to the alternative models, in particular over the longer forecast horizons. The mean model fares very poorly in these comparisons, which is consistent with its difficulties predicting, among other variables, the short-term nominal interest

[3] The (log) determinant statistic of an expanded MSE-matrix based on the forecast errors for the 1-step ahead until the H-steps ahead forecasts is, as pointed out by Clements and Hendry, invariant to these transformations, but due to short forecast samples it is often not possible to calculate such a measure in practice. Moreover, it should be kept in mind that their measure need not be invariant to increases in the maximum forecast horizon, H.

rate. Turning to the trace statistic when forecasting annual changes of the variables in first differences, the forecasting performance of the NAWM is, however, less impressive; see Figure 7. The main reason for this is undoubtedly the reweighing of the underlying innovations, where forecast errors prior to one year ahead when forecasting quarterly changes now matter for the forecasts of annual changes beyond one year.

It should be noted that the chosen scaling matrix does not work in favour of the NAWM when using the trace statistic. When $M = I$, the NAWM is very competitive also at the shorter horizons when forecasting quarterly changes and is often among the best at all horizons for the annual changes. The primary explanation for this is that the NAWM forecasts the more volatile variables (trade variables, short-term nominal interest rate, and the real effective exchange rate) relatively well, while the variables where it is less successful (nominal wage growth, real private consumption growth, and consumption deflator inflation) are less volatile.

The log determinant statistic is invariant across forecasting models to the choice of scaling matrix and is displayed in Figures 8 and 9 when forecasting quarterly and annual changes, respectively. For quarterly changes we find that the NAWM is competitive at the longer forecast horizons in all 3 cases. Overall, the large BVAR with a mixed prior tends to outperform the other models, especially at the shorter horizons. When forecasting annual changes, the log determinant statistic again tends to favour the large BVAR with a mixed prior, but now the random-walk model performs almost as well in the 7 and 3 variable cases. Moreover, at longer forecast horizons the mean model often performs well, in particular for the 12 and 7 variable cases.

Since the trace (log determinant) is equal to the sum of (the log of) the eigenvalues, it is possible to make additional interpretations of the multivariate point forecast results by performing a singular value decomposition of the MSE-matrices and compute decompositions of the MSEs based on the shares due to the different eigenvalues. That is, let $\Sigma_M(h) = V\Lambda V'$, where V is the matrix with eigenvectors with typical element v_{ij}, $V'V = I_s$, while $\Lambda = \text{diag}[\lambda_1, \ldots, \lambda_s]$ is a diagonal matrix with the eigenvalues in descending order. It now follows that the share of the h-steps ahead forecast error variance of variable i due to eigenvalue j is given by:

$$\sigma_{ij,M}(h) = \frac{v_{ij}^2 \lambda_j}{\sum_{j=1}^{s} v_{ij}^2 \lambda_j}, \quad i, j = 1, \ldots, s.$$

Since large (small) eigenvalues are equal to the least (most) predictable dimensions at a given forecast horizon, MSE-based variance decompositions may help us link the larger (smaller) eigenvalues to certain variables.

For the NAWM we find that the largest eigenvalue always explains most of the forecast error variance of nominal wage growth in the 12 and 7 variable cases. At the same time, the smaller eigenvalues typically explain a large share of the forecast error variance in employment and to a lesser extent in the short-term nominal interest rate. Hence, the NAWM is typically punished for its poor performance when forecasting nominal wage growth and using the trace statistic, while its relatively good performance from the perspective of the log determinant is to a fairly large

extent due to its employment forecasts. For the case with 3 variables, the largest eigenvalue is similarly linked with GDP deflator inflation and the smallest with the short-term nominal interest rate. While the RMSEs in Figures 4 and 5 suggest similar interpretations, it should be kept in mind that, unlike the MSE-based decompositions, they do not take the correlation structure into account and need therefore not be consistent with the eigenvalue-based forecast error variance decompositions.

5.2. Density Forecasts

Dawid (1984) pointed out that an important purpose of statistical analysis is to not only make sequential forecasts of the future, but also to provide suitable measures of the uncertainty that is linked to them. That is, forecasts are both probabilistic and sequential in nature, taking the form of probability distributions over a sequence of future values. This basic premise is the foundation of what Dawid called the *prequential* approach.

While point forecasts are sometimes of first-order importance, forecast uncertainty has since Dawid's article been given an increasingly more important role with both important methodological developments (see, e.g., Diebold, Gunther, and Tay, 1998; Christoffersen, 1998; Amisano and Giacomini, 2007) and interesting empirical applications (Diebold, Tay, and Wallis, 1999; Clements and Smith, 2000; Adolfson, Lindé, and Villani, 2007); see also Tay and Wallis (2000) for a survey. Moreover, the use of uncertainty bands in the inflation reports of several central banks (e.g., the Bank of England and Sveriges Riksbank) has become instrumental in communicating with the public.

The predictive density makes it feasible to take forecasting uncertainty into account and may also be used to evaluate the goodness-of-fit of a model. It is well known (see, e.g., Section 2.6.2. in Geweke, 2005) that the predictive density can be expressed as the ratio between the marginal likelihood of a model for the extended sample $\mathcal{Y}_{T+H}$ and the estimation sample $\mathcal{Y}_T$. That is,

$$p\big(y_{T+1},\dots,y_{T+H}|\mathcal{Y}_T,m\big)=\frac{p\big(\mathcal{Y}_{T+H}|m\big)}{p\big(\mathcal{Y}_T|m\big)}.$$

Hence, the height of the predictive density at any realised values $y_{T+1},\dots,y_{T+H}$ is equal to the improvement in the marginal likelihood when the extended sample includes these future values.

Provided that the models we wish to compare are not subject to Lindley's paradox (Bartlett, 1957; Lindley, 1957), the marginal likelihood is one important criterion for evaluating the goodness-of-fit of a model estimated with Bayesian methods: the larger the value of the marginal likelihood, the better a model fits the observed data. Since the height of the predictive density is equal to the ratio of the marginal likelihood for the extended sample and the estimation sample, model m_i, say, may have a larger value for the height of the predictive density than model m_j, but still have lower values for the marginal likelihood than model m_j for both the full sample and the estimation sample.

Nevertheless, if the focus of our attention is out-of-sample forecasting, then the within-sample goodness-of-fit or lack thereof should not directly be a concern. Moreover, as Geweke and Amisano (2009) and Gneiting, Balabdaoui, and Raftery (2007) note, the assessment of a predictive distribution on the basis of its density and the observed data only is consistent with the prequential approach.

Scoring rules are used to evaluate the quality of probabilistic forecasts by giving a numerical value using the predictive distribution and an event or value that materialises. A scoring rule is said to be *proper* if the forecaster maximises the expected score (utility) for an observation drawn from a distribution D_i when the forecaster gives the probabilistic forecast D_i rather than $D_j \neq D_i$. Furthermore, a scoring rule is said to be *strictly proper* if the maximum is unique. Proper scoring rules are therefore important since they encourage the forecaster to be honest; i.e., there is no gain from reporting D_j instead of D_i.

A widely used scoring rule that was suggested by Good (1952) is the log predictive score. We may define this scoring rule by:

$$S(m) = \frac{1}{N} \sum_{t=T}^{T+N-1} \log p\big(y_{t+1}, \ldots, y_{t+H} | \mathcal{Y}_t, m\big). \tag{21}$$

If the value of the predictive density only depends on the actual realisations of y over the prediction sample, 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 observed data. For a survey of scoring rules, see Gneiting and Raftery (2007).

We can similarly define a log scoring rule within the parametric classical framework. Let $\hat{\theta}_m^t$ be the maximum likelihood estimator of θ_m using the data $\mathcal{Y}_t$ and the likelihood $p(\mathcal{Y}_t; \theta_m, m)$. The log predictive score is now given by:

$$S(m) = \frac{1}{N} \sum_{t=T}^{T+N-1} \log p\big(y_{t+1}, \ldots, y_{t+H} | \mathcal{Y}_t, \hat{\theta}_m^t, m\big), \tag{22}$$

where $p(y_{t+1}, \ldots, y_{t+H} | \mathcal{Y}_t, \hat{\theta}_m^t, m)$ is the predictive likelihood.

When evaluating the density forecast of the NAWM and of the alternative forecasting models we shall focus on the h-steps ahead forecasts. This means that we consider the log predictive score function

$$S_h(m) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \log p\big(y_{t+h} | \mathcal{Y}_t, m\big), \quad h = 1, \ldots, H, \tag{23}$$

when using Bayesian methods. For the models which are estimated with classical methods, we use an expression such as in (22), but where the likelihood function concerns $y_{t+h}|\mathcal{Y}_t$ and N is replaced with N_h.

The relationship between the marginal likelihood and the log score function in (23) holds when $h = 1$, but breaks down for $h > 1$. As pointed out by Adolfson, Lindé, and Villani (2007), this means that the marginal likelihood cannot be used to assess whether or not some model

performs well on certain forecast horizons, while other models do better on other horizons. Moreover, to compute $S_h(m)$ in (23) for $h > 1$ is generally not a simple matter for models estimated with Bayesian methods since $p\big(y_{t+h}|\mathcal{Y}_t, m\big)$ does not have a known analytical form. Furthermore, estimation of the density via kernel density estimation techniques is not practical when the dimension, n, is large. Following Adolfson, Lindé, and Villani (2007) we shall therefore approximate the predictive density with the multivariate normal.

The mean and the covariance matrix for the h-steps ahead forecasts are calculated for each t in the forecast sample using the 250,000 simulated values ($M_1 = M_2 = 500$) of y_{t+h} for the DSGE model and the two large BVARs with $n = 18$. Similarly, the mean and the covariance matrix for the small BVAR with a steady-state prior are computed from the 250,000 h-steps ahead forecast values per t of z_{t+h}. For the small VAR model, the predictive likelihood is multivariate normal with mean equal to $z_{t+h|t} = E[z_{t+h}|\mathcal{Y}_t, \theta_m, m]$ and covariance matrix $\Sigma = E[(z_{t+h} - z_{t+h|t})(z_{t+h} - z_{t+h|t})'; \theta_m, m]$, evaluated at the maximum likelihood estimates based on the conditioning information.

In addition, we can compute the log predictive score for the random-walk and mean models by assuming that the 1-step ahead prediction errors are multivariate normal with mean zero and covariance matrix Σ. The 1-step ahead population covariance matrix is estimated using the conditioning information. For the random-walk model we have that the h-steps ahead population forecast error covariance matrix is equal to $h\Sigma$, while it is equal to Σ for all h in the mean model.

The density forecast evaluation will again focus on the three cases with 12, 7, and 3 variables that we studied in the previous section. The assumption that the predictive density can be approximated by a multivariate normal is particularly convenient since we may simply use the properties that (i) the distribution of any subset of the variables is also normal, and (ii) the mean and the covariance matrix of the distribution for the subset is equal to the same subset of the mean and the covariance matrix of the joint distribution.[4]

The log predictive scores when forecasting quarterly changes of the variables in first differences are shown in Figure 10. It is striking that for all three cases and all forecast horizons, the large BVAR with a mixed prior obtains the highest value for the log predictive score. At the other end of the spectrum, we find that for the cases where the small VAR model can be evaluated, it always obtains the lowest value. The other 5 models therefore rank somewhere in between, with the DSGE model often coming close to the large BVAR with a mixed prior, especially at the longer horizons and in the 3 variable case. The random-walk model is here ranked close to the DSGE model for the shorter but not the longer horizons. The same can be said for the large BVAR model with a white-noise prior, while the small BVAR with a steady-state prior also tends to be more competitive relative to the DSGE model at the longer horizons.

[4] In the case of the DSGE model, the normality assumption is probably not so critical since most of the forecast error variance is explained by factors other than parameter uncertainty; see, e.g., Adolfson, Lindé, and Villani (2007, Figure 4). Similar results are available for the NAWM. For the VAR models, however, parameter uncertainty is considerably more important and, hence, the assumption of normality is more likely to be questionable.

In Figure 11 we find the log predictive scores when forecasting annual changes of the variables in first differences. It was pointed out in Section 5.1 that the multivariate MSE-based trace and log determinant statistics are not invariant to non-singular, scale-preserving linear transformations in linear models. The log predictive score in (23) is similarly not invariant to the choice of predicting quarterly or annual changes of the variables other than for 1-step ahead forecasts.[5] For example, the small BVAR with a steady-state prior ranks below the DSGE when $h = 7$ when forecasting quarterly changes, and above the DSGE at the same horizon when forecasting annual changes. Nevertheless, the rankings of the DSGE and the different VAR models are very stable when comparing the quarterly to the annual changes.

Regarding the mean and the random-walk models, the annual and quarterly models are not isomorphic and, hence, we do not expect any forecast evaluation criteria to preserve their ranks. For example, the random-walk model is very competitive when forecasting annual changes, where the value of the log predictive score is often close to that of the large BVAR with the mixed prior, and is always better than the DSGE model. By contrast, the random-walk model when forecasting quarterly changes is often ranked towards the bottom, as in the 12 variable case. Still, it should be borne in mind that the h-steps ahead covariance matrix of the random-walk and mean models do not reflect parameter uncertainty. Hence, the rankings of these models may be boosted by the choice of covariance estimator.

5.3. Relating the forecast performance of the DSGE model to its structure

We identify two main factors which explain the relative strengths and weaknesses of the NAWM in the forecasting exercise. On the one hand, the NAWM builds on explicit micro-foundations which allows to derive a parsimoniously parameterised structure respecting a large number of cross-equation restrictions. On the other hand, because of the assumed balanced-growth path, the NAWM's flexibility to deal with differing trends in the data is rather limited, when compared to the reduced-form models. While parsimony is likely to be an advantage for achieving forecast accuracy, an excessively rigid treatment of trends may give rise to a bias in the forecasts and, hence, inflate the RMSEs.

Concerning the role of trends, the NAWM's balanced-growth path implies tight restrictions on the mean of the growth rates of its observed variables. First, the model assumes common growth rates for the subsets of real and nominal variables, respectively. Second, these growth rates are constant over time and there is no further updating in a Bayesian sense. For example, the deterministic steady-state component of the common growth rate for the real variables is calibrated to equal 2 percent per annum, which is close to the average of annual real GDP growth over the full estimation sample. Medium-run deviations from this deterministic growth component are captured by the model's permanent technology shock. A positive permanent

[5] If we compute the log predictive score using (21), the ranking between models is invariant to scale-preserving linear transformations of the variables. Still, the ranking may change when extending the maximum forecast horizon from H to, say, $H + 1$.

technology shock implies a permanent increase in the levels and a transitory increase in the growth rates of the real variables with a half-life of 3 quarters. All other shocks display a rather fast mean reversion in terms of growth rates, implying a strong role for the deterministic steady-state growth rate for the higher forecast horizons.

Maintaining a common, and constant, steady-state growth rate for groups of variables is inducing two types of biases in the forecasts, as can be inferred from Table 1. First, the steady-state growth rate of some variables differs from the mean growth rate over the forecast evaluation sample. For example, the NAWM assumes that, in steady state, investment grows at the common growth rate of 2 percent per annum, while observed investment growth is actually 2.7 percent per annum. Consequently, the model tends to underpredict investment growth. Second, there are notable differences between the mean growth rates of the different variables within a specific group. For example, in contrast to investment growth, private consumption growth over the forecast evaluation period has been below the model's steady-state growth rate of 2 percent. Furthermore the NAWM's cross-equation restrictions imply that a bias in the forecast for one variable might be transmitted through the model.

To investigate in some more detail the most problematic dimension for the NAWM's nominal variables, we have plotted the 1 to 8-steps ahead mean forecast paths of quarterly nominal wage growth for all forecasting models in the upper part of Figure 12. The NAWM generally overpredicts nominal wage growth in a manner that resembles the behaviour of the forecasts from the mean model, as can be seen from Figure 4. We can relate this overprediction to the difference between the steady-state growth rate of real wages of 1.2 percent per annum and the observed mean growth over the forecast sample which is only 0.3 percent per annum. Over the prediction horizon the point forecasts of real wages are returning to the model's steady state. In terms of observable variables this implies that nominal wage growth is overpredicted and private consumption deflator inflation is underpredicted. The systematic overprediction is also affecting the expectations of households which are repeatedly expecting higher real wage growth. This in turn implies an overprediction of real private consumption. Accordingly, the predictions for nominal wages, the private consumption and the GDP deflator, as well as real private consumption tend to have relatively large mean errors, as can be seen in Table 2.

The fluctuations of the NAWM's variables around the balanced-growth path respect a tightly specified economic structure, based on intertemporal optimality conditions. From the equations reported in Section 2 it is apparent that the parameterisation of the model is very parsimonious. Taking the consumption equation (1) as an example, we can see that this equation depends on one estimated parameter κ (the habit formation parameter) and one calibrated parameter g_z (the deterministic steady-state trend growth rate of productivity). Since all equations are solved simultaneously we can derive a reduced-form representation of the model (equations (10) and (11)) which obeys the underlying cross-equation restrictions. This reduced-form representation

depends on 45 estimated parameters and explains the dynamics of the model's 12 observed endogenous variables.

In comparison to a VAR the number of parameters is significantly reduced. This economically motivated shrinkage, in combination with the implied cross-equation restrictions, is likely to improve the forecasting performance. Sims (1980) argued that most of the restrictions in macroeconomic models are false, but the models might still be useful tools for forecasting and policy analysis unless the restrictions are "very false". Moreover, employing Bayesian inference methods, which combine prior information obtained from earlier studies at both the micro and the macro level with the likelihood function for the data, results in rather tightly estimated structural parameters, which account for only a small part of the dispersion of the NAWM's predictive distributions. This compares with a relatively high share of parameter uncertainty in the predictive uncertainty of the BVARs.

The advantages of the parsimonious and tight parameterisation of the NAWM can be inferred, for example, from the lower part of Figure 12 which depicts the 1 to 8-steps ahead mean forecast paths for the short-term nominal interest rate across models. Compared with the reduced-form models, the NAWM fares very well, consistent with the pattern of the RMSEs shown in Figure 4.

6. Summary and Conclusions

In a thought-provoking article in the *Journal of Economic Perspectives*, Diebold (1998) speculated on the future of macroeconomic forecasting. Unlike many other observers at that time, he did not view the failure of the large-scale system-of-equations macroeconomic forecasting models (that had been popular until the 1970s) as signifying a bleak future for macroeconomic forecasting. From what began with neoclassical models under the rational expectations paradigm, such as the real business cycle model in Kydland and Prescott (1982) or linear-quadratic models as in Hansen and Sargent (1980), and nonstructural time-series models like the VAR in Sims (1980), Diebold (1998, p. 189) predicted that a hallmark of macroeconomic forecasting in the first 20 years of the 21st century would be:

> ... a marriage of the best of the nonstructural and structural approaches, facilitated by advances in numerical and simulation techniques that will help macroeconomists to solve, estimate, simulate, and yes, *forecast* with rich models.

The highly influential studies by Christiano, Eichenbaum, and Evans (2005), Smets and Wouters (2003, 2007), and Adolfson, Laséen, Lindé, and Villani (2007) have contributed greatly to the development of DSGE models since the time when Diebold made his predictions, while Del Negro and Schorfheide (2004, 2006) and Del Negro, Schorfheide, Smets, and Wouters (2007) are important examples of what may be regarded as a "marriage" of the nonstructural and structural approaches that Diebold referred to. Nevertheless, among the many insightful predictions made by Diebold, one factor that seems to have surpassed his expectations concerns the potential increase in the scale of DSGE models. Where Diebold viewed it possible to see no more than eight

or ten variables in equilibrium, the DSGE model in Adolfson, Laséen, Lindé, and Villani (2007), based on euro area data, has 12 of its 15 observed variables that are endogenously determined by the model, while the 3 foreign variables are exogenous to the rest of the system and modelled as a structural VAR. Similarly, 12 of the 18 observed variables in the NAWM are endogenously determined by the internal mechanisms of the DSGE model, while Christiano, Trabandt, and Walentin (2009) have extended this dimension even further when applying their DSGE model to Swedish data.

In this paper we have reviewed forecasting with DSGE models, using the NAWM as an illustration. This DSGE model was designed for regular use at the euro area level in the macroeconomic projections undertaken by ECB/Eurosystem staff. The forecast evaluation exercise that we have conducted covers both point and density forecasts. As a consequence, we have discussed estimation of the predictive distribution of a DSGE model based on Bayesian methods, as well as the estimation of moments of the marginal h-steps ahead distributions. We have also discussed relevant benchmarks for the DSGE model, such as forecasts taken from VARs, BVARs, a random walk, and a location parameter, namely the mean.

The out-of-sample forecast evaluation exercise covers the period after the introduction of the euro and focuses on the 12 observed variables in the NAWM that are endogenously determined by the model. Overall, the results suggest that the NAWM performs quite well when compared with the reduced-form forecasting tools. In particular, the model compares favourably when forecasting real GDP growth, the trade variables, employment, the real exchange rate, and the short-term nominal interest rate. However, the NAWM is less successful when forecasting certain nominal variables, in particular nominal wage growth. One explanation for this is that the year-on-year steady-state growth of nominal wages is 3.1 percent in the NAWM, while wage moderation over the forecast evaluation period has kept nominal wage growth down at around 2.3 percent. The relatively strong mean reversion properties of the model therefore lead to persistent negative forecast errors.

Nevertheless, the results in this paper support earlier studies of the forecasting ability of DSGE models. At this stage of their development, they can compete when we use out-of-sample forecast performance as a measure of fit. Naturally, this does not mean that they necessarily "win" the competition in all dimensions. Moreover, Clements and Hendry (2005) emphasise that forecast performance is not a good instrument for evaluating models in general, except when the model is intended for forecasting; see also Granger (1999). In particular, they note that a "good" out-of-sample forecast performance should not be viewed as a "seal of approval" to the model or the theory it may be based on. Similarly, poor performance need not imply that the model or the theory is invalidated.

Still, the forecasting performance of the NAWM in this study, as well as the performance of DSGE models documented in previous studies, is quite impressive, not least in view of the large number of cross-equation restrictions that are imposed. Moreover, it is important to recall that

forecasting (and policy analysis) with false restrictions may not hurt the performance of a model. In fact, as long as the restrictions are not "very false", they may even help a model to function for these purposes; see, e.g., Sims (1980, Section 1D). A DSGE model—like all macroeconomic models—is a simplification of an actual economy and is therefore, one may argue, misspecified. The degree to which such misspecification matters for, say, policy analysis may be diagnosed by making use of tools that allow us to study departures from the restrictions implied by the model. With the aid of one such tool, DSGE-VARs, Del Negro, Schorfheide, Smets, and Wouters (2007) note that misspecification of the DSGE model they estimate is not so large that it prevents its use in policy analysis, but that it remains a concern.

Appendix A: The BVAR with a Steady-State Prior

Following Villani (2009) we assume that Ψ is a priori independent of Π_l and Ω with $\text{vec}(\Psi) \sim N(\mu_\psi, \Sigma_\psi)$ and Σ_ψ being positive definite. Regarding the parameters on lags of the endogenous variables, we define $\Pi = [\Pi_1 \ \cdots \ \Pi_k]$ and assume that $\text{vec}(\Pi) \sim N(\mu_\pi, \Sigma_\pi)$. Finally, we use a diffuse prior on Ω, as represented by the well-known form $p(\Omega) \propto |\Omega|^{-(p+1)/2}$.

To parameterise the prior on Π we assume that the prior mean of Π_l is zero for all $l \geq 2$. For the first lag all off-diagonal elements are assumed to be zero, while the diagonal elements are equal to λ_D when $z_{i,t}$ is a first differenced variable (e.g., GDP growth), and given by λ_L when $z_{i,t}$ is a level variable (e.g., the nominal interest rate). Regarding the parameterisation of Σ_π we use a Minnesota-style prior; cf. Doan, Litterman, and Sims (1984); Litterman (1986). Letting $\Pi_{ij,l}$ denote the element in row (equation) i and column (on variable) j for lag l. The matrix Σ_π is assumed to be diagonal with

$$\text{Var}\big(\Pi_{ij,l}\big) = \begin{cases} \dfrac{\lambda_o^2}{l^{\lambda_h}}, & \text{if } i = j, \\ \dfrac{\lambda_o^2 \lambda_c \Omega_{ii}}{l^{\lambda_h} \Omega_{jj}}, & \text{otherwise.} \end{cases} \tag{A.1}$$

The parameter Ω_{ii} is simply the variance of the residual in equation i and, hence, the ratio Ω_{ii}/Ω_{jj} takes into account that variable i and variable j may have different scales.

Formally, this parameterisation is inconsistent with the prior being a marginal distribution since it depends on Ω. As is common for the Minnesota type of prior we deal with this by replacing the Ω_{ii} parameters with the within-sample maximum likelihood estimate. The hyperparameter $\lambda_o > 0$ gives the overall tightness of the prior around the mean, while $0 < \lambda_c < 1$ is the cross-equation tightness hyperparameter. Finally, the hyperparameter $\lambda_h > 0$ measures the harmonic lag decay.

In the empirical application the BVAR model has 7 variables that are taken from the observed variable set for the NAWM. The variables we have selected are the same type of variables as were used by Smets and Wouters (2003). They are: real GDP growth, real private consumption growth, real total investment growth, GDP deflator inflation, employment, nominal wage growth, and the short-term nominal interest rate. For the steady-state prior we let:

$$\mu_\psi' = \begin{bmatrix} 0.5 & 0.5 & 0.5 & 0.475 & 0 & 0.775 & 4.4 \end{bmatrix},$$
$$\text{diag}(\Sigma_\psi)' = \begin{bmatrix} 1 & 1 & 4 & 1 & 0.5 & 1 & 5 \end{bmatrix},$$

while all off-diagonal elements of Σ_ψ are zero. The hyperparameters for the Π_l parameters are given by $\lambda_L = 0.9$, $\lambda_D = 0$, $\lambda_o^2 = \lambda_c = 0.5$, while $\lambda_h = 1$. All variables are treated as first differenced variables except employment and the short-term nominal interest rate. The lag order, k, is set to 4.

Appendix B: Dummy Observation Prior for the Large BVAR Models

The VAR model in (19) can be rewritten more compactly as:

$$y_t = \beta x_t + \varepsilon_t, \tag{B.1}$$

where $\beta = [\Pi\ \Phi]$ is an $n \times (nk+1)$ matrix, and $x_t = [y'_{t-1} \cdots y'_{t-k}\ 1]'$. The conjugate normal/inverted Wishart prior is represented by the pair $\text{vec}(\beta)|\Omega \sim N(\mu_\beta, \Omega_\beta \otimes \Omega)$ and $\Omega \sim IW(A, v)$, where $\otimes$ is the Kronecker product. By constructing $T_d = n(k+1)+1$ dummy observations for y_t and x_t, collected in the matrices $y_{(d)}$ $(n \times n(k+1)+1)$ and $x_{(d)}$ $(nk+1 \times n(k+1)+1)$, we can link these matrices to the prior hyperparameters of the Minnesota prior; see Appendix A. Specifically, $\Omega_\beta = (x_{(d)}x'_{(d)})^{-1}$, $\mu_\beta = \text{vec}(\beta_0)$, $\beta_0 = y_{(d)}x'_{(d)}\Omega_\beta$, $A = (y_{(d)} - \beta_0 x_{(d)})(y_{(d)} - \beta_0 x_{(d)})'$, while $v = T_d - (nk+1) + 2$ so that the prior mean of Ω exists.

To ensure that the Kronecker structure in the prior on β is feasible, the cross-equation tightness parameter of the Minnesota prior is set to unity, i.e., $\lambda_c = 1$. In addition, the harmonic lag decay hyperparameter is given by $\lambda_h = 2$. Next, let ω_i be the scale parameter of residual i, while δ_i is the prior mean of the diagonal element i in Π_1. The dummy observation matrices are now:

$$\begin{aligned} y_{(d)} &= \begin{bmatrix} (1/\lambda_o)\text{diag}[\delta_1\omega_1, \ldots, \delta_n\omega_n] & 0_{n\times n(k-1)} & \text{diag}[\omega_1, \ldots, \omega_n] & 0_{n\times 1} \end{bmatrix} \\ x_{(d)} &= \begin{bmatrix} (1/\lambda_o)\big(J_k \otimes \text{diag}[\omega_1, \ldots, \omega_n]\big) & 0_{nk\times n} & 0_{nk\times 1} \\ 0_{1\times nk} & 0_{1\times n} & \epsilon \end{bmatrix}, \end{aligned} \tag{B.2}$$

where $J_k = \text{diag}[1, \ldots, k]$, while ϵ is a very small number which handles the use of an improper prior on Φ.

The sum of the Π_l matrices part of the prior is implemented by appending the n dummy observations $(1/\tau)\text{diag}[\delta_1\mu_1, \ldots, \delta_n\mu_n]$ to the $y_{(d)}$ matrix, and $[(1/\tau)(\imath'_k \otimes \text{diag}[\mu_1, \ldots, \mu_n])\ 0_{n\times 1}]'$ to the $x_{(d)}$ matrix. The hyperparameter $\tau > 0$ takes care of shrinkage, where $\tau \to 0$ means that the prior on $(I_n - \sum_{l=1}^{k} \Pi_l)$ approaches the case of exact differences, while shrinkage decreases as τ becomes larger. The hyperparameter μ_i reflects the mean of y_{it}, while $\imath_k$ is a $k \times 1$ unit vector. The total number of dummy observations is therefore $T_d = n(k+2)+1$.

In the empirical applications, $\tau = 10\lambda_o$, i.e., a relatively loose prior on the sum of the autoregresive matrices. For the BVAR with a white-noise prior, $\delta_i = 0$ for all variables, while the BVAR with a mixed prior has $\delta_i = 0$ if y_{it} is a first differenced variable and $\delta_i = 1$ when y_{it} is a level variable. The ω_i hyperparameter is given by the within-sample residual standard deviation from an AR(k) model for y_{it}, while μ_i is given by the within-sample mean of y_{it}. The lag order is $k = 4$.

The formula suggested by Bańbura, Giannone, and Reichlin for selecting λ_o can here be expressed as

$$\lambda_o(\phi) = \arg\min_\lambda \left| \phi - \frac{1}{q}\sum_{j=1}^{q} \frac{\sigma_j^2(\lambda)}{\sigma_j^2(0)} \right|, \tag{B.3}$$

where $\phi \in (0,1)$ is the desired fit, and $\sigma_i^2(\lambda)$ is the 1-step ahead mean square forecast error of variable i when $\lambda_o = \lambda$. The 1-step ahead within-sample mean square forecast errors used in the selection scheme are based on the sample 1985Q1-1998Q4. With $\phi = 0.5$, $q = 3$ using real GDP growth, GDP deflator inflation and the short-term nominal interest rate, this selection scheme gives $\lambda_o = 0.0827$ for the white-noise prior, and $\lambda_o = 0.0693$ for the mixed prior.

With $y = [y_{(d)}\ y_1\ \cdots\ y_T]$ and $x = [x_{(d)}\ x_1\ \cdots\ x_T]$, the joint posterior distribution of (β, Ω) is given by the pair:

$$\text{vec}(\beta)|\Omega, \mathcal{Y}_T \sim N\big(\bar{\mu}_\beta, (xx')^{-1} \otimes \Omega\big), \quad \Omega|\mathcal{Y}_T \sim IW(\bar{\Omega}, T + T_d + 2 - (nk+1)), \tag{B.4}$$

where $\bar{\mu}_\beta = \text{vec}(\bar{\beta})$, $\bar{\beta} = yx'(xx')^{-1}$, and $\bar{\Omega} = (y - \bar{\beta}x)(y - \bar{\beta}x)'$. The marginal posterior distribution of β is matrix-t; see, e.g., Bauwens, Lubrano, and Richard (1999, Theorem A.19).

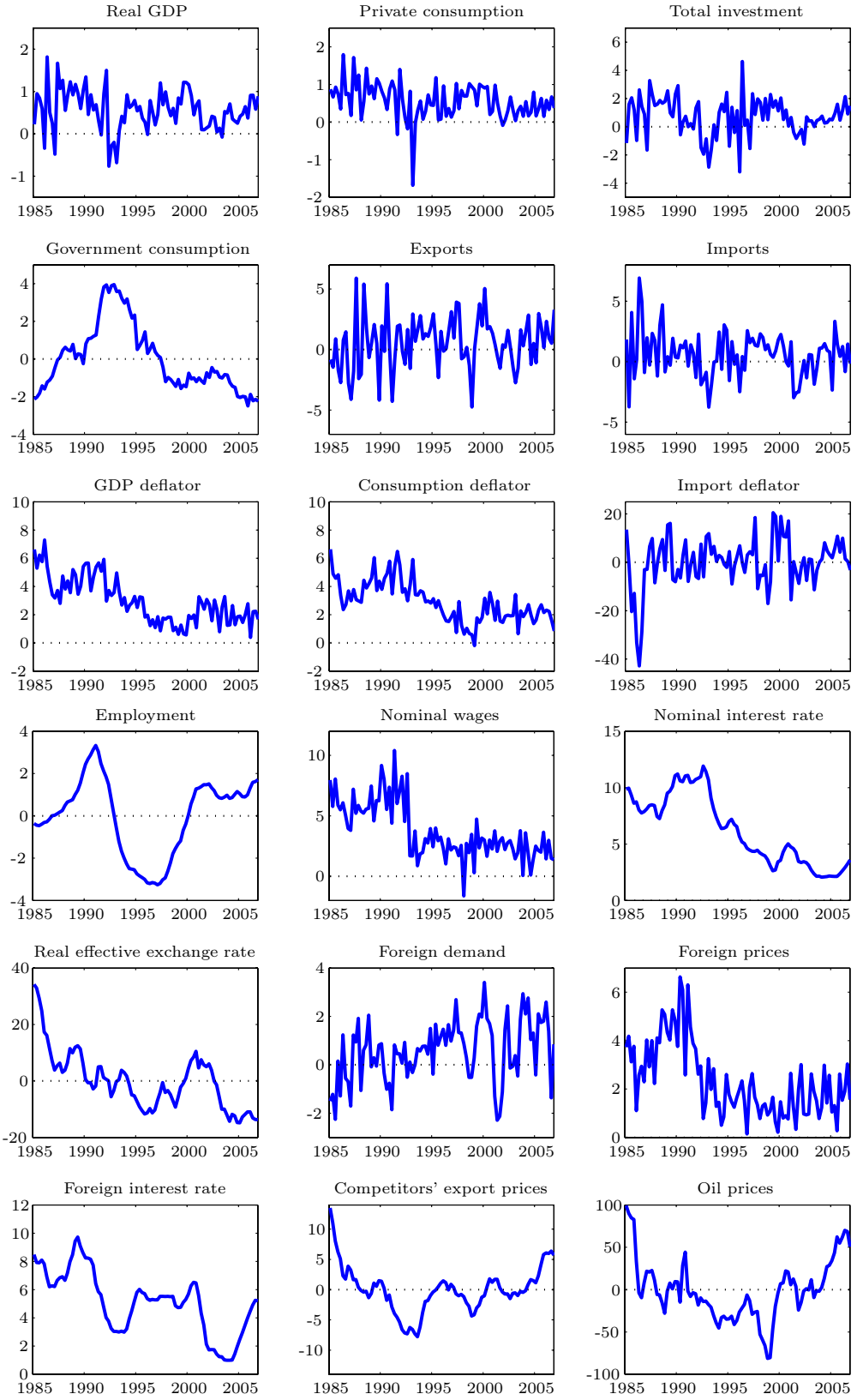
TABLE 1. Steady-state values of the NAWM and sub-sample means of selected variables.

Variable	Steady state	1985Q1-2006Q4	1985Q1-1998Q4	1999Q1-2006Q4
Real GDP	0.500	0.568	0.586	0.537
Private consumption	0.500	0.554	0.614	0.450
Total investment	0.500	0.708	0.729	0.671
Exports	0.500	0.568	0.378	0.901
Imports	0.500	0.568	0.774	0.208
GDP deflator	0.475	0.749	0.906	0.475
Consumption deflator	0.475	0.725	0.854	0.500
Import deflator	0.475	0.009	−0.386	0.701
Employment	0.000	0.000	−0.486	0.851
Nominal wages	0.775	0.933	1.135	0.581
Nominal interest rate	4.400	6.225	8.020	3.082
Real exchange rate	0.000	0.000	2.485	−4.349

TABLE 2. Percentage share for squared mean errors of mean squared errors when forecasting quarterly changes of variables in first differences.

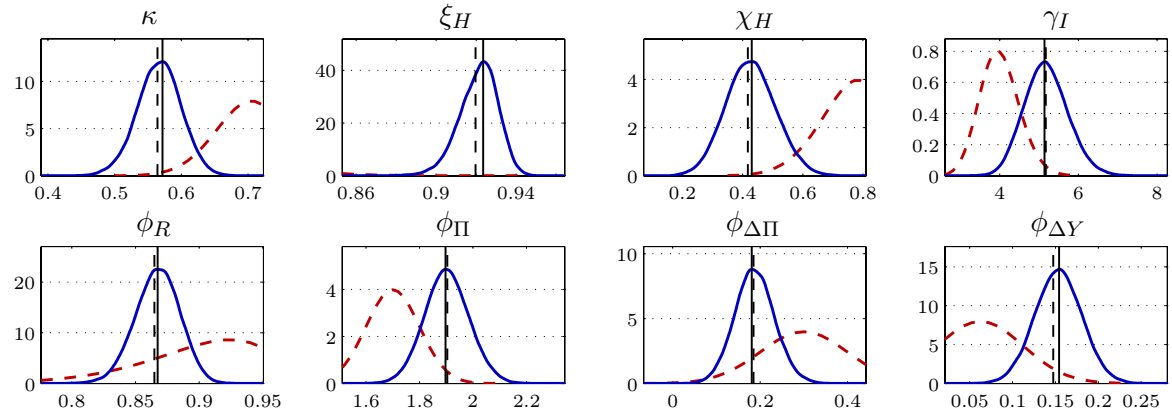
	NAWM			BVAR - mixed prior			Mean		
Variable	$h=1$	$h=4$	$h=8$	$h=1$	$h=4$	$h=8$	$h=1$	$h=4$	$h=8$
Real GDP	0.0	28.9	32.7	0.2	0.4	0.6	2.0	6.8	24.5
Private consumption	12.6	62.9	51.7	11.2	7.8	4.4	18.6	27.5	46.6
Total investment	39.1	34.1	0.2	0.0	0.1	0.2	0.4	4.7	12.4
Exports	0.5	0.8	1.3	4.9	1.6	0.6	4.9	2.8	0.0
Imports	0.2	0.0	8.3	2.5	0.6	0.2	7.8	8.6	15.3
GDP deflator	8.3	43.6	32.7	10.6	8.6	4.1	70.4	71.6	72.3
Consumption deflator	21.9	65.1	45.5	5.4	3.0	8.7	67.1	71.6	78.2
Import deflator	1.8	3.8	2.1	0.3	0.0	3.6	12.0	9.3	3.2
Employment	7.5	28.4	32.8	1.9	0.0	0.5	73.8	91.1	95.8
Nominal wages	45.1	79.2	78.8	0.1	25.7	16.0	75.0	80.5	81.0
Nominal interest rate	25.7	2.3	12.7	54.8	0.9	0.1	96.5	96.9	97.3
Real exchange rate	1.2	5.5	17.6	3.5	11.3	23.1	38.7	37.3	49.5

FIGURE 1. The Data.



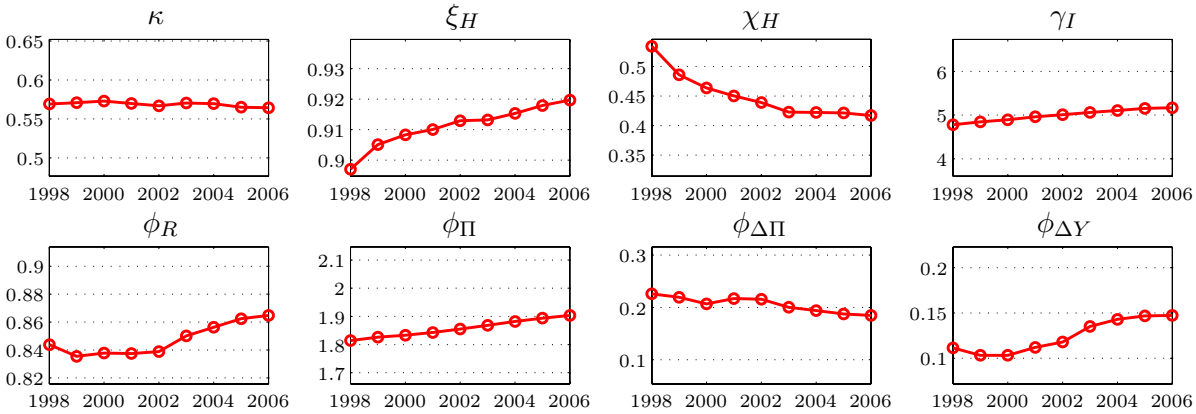
Note: This figure shows the time series of the observed variables used in the estimation of the NAWM. Details on the variable transformations are provided in Section 3.2. Inflation and interest rates are reported in annualised percentage terms.

FIGURE 2. Prior and posterior densities of selected structural parameters of the NAWM.



Note: The marginal posterior densities are based on 550,000 draws (blue solid line) and are plotted against their marginal prior densities (red dotted line), with 50,000 draws being discarded as burn-in sample. The solid vertical black line is the marginal mode and the dashed vertical black line the joint mode.

FIGURE 3. Recursive posterior mode estimates of selected structural parameters of the NAWM.



Note: The parameters have been estimated recursively with the estimation sample being gradually extended by a full year from 1998Q4 to 2006Q4.

FIGURE 4. Scaled root mean squared forecast errors for 12 variables when forecasting quarterly changes of variables in first differences.

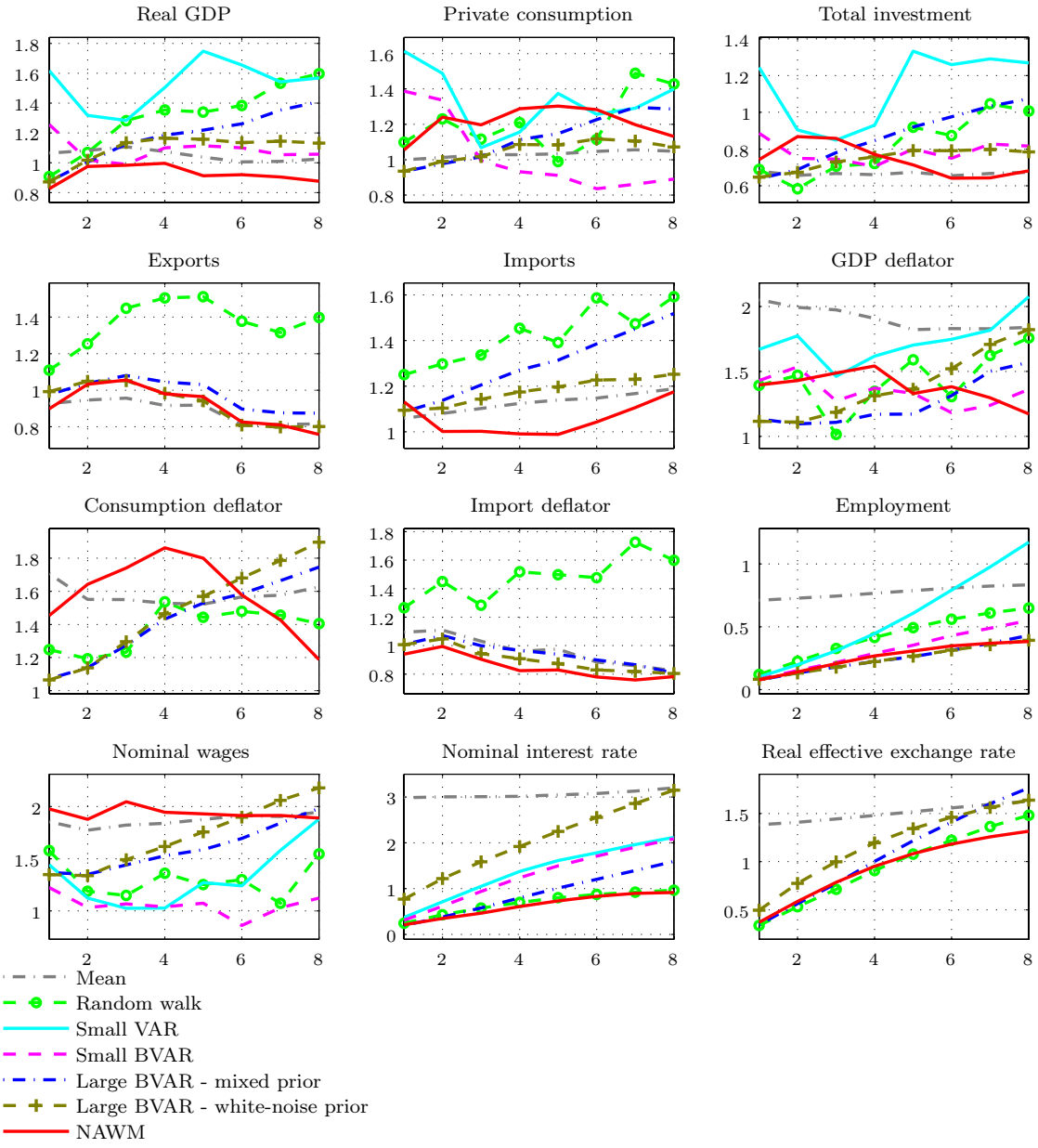


FIGURE 5. Scaled root mean squared forecast errors for 12 variables when forecasting annual changes of variables in first differences.

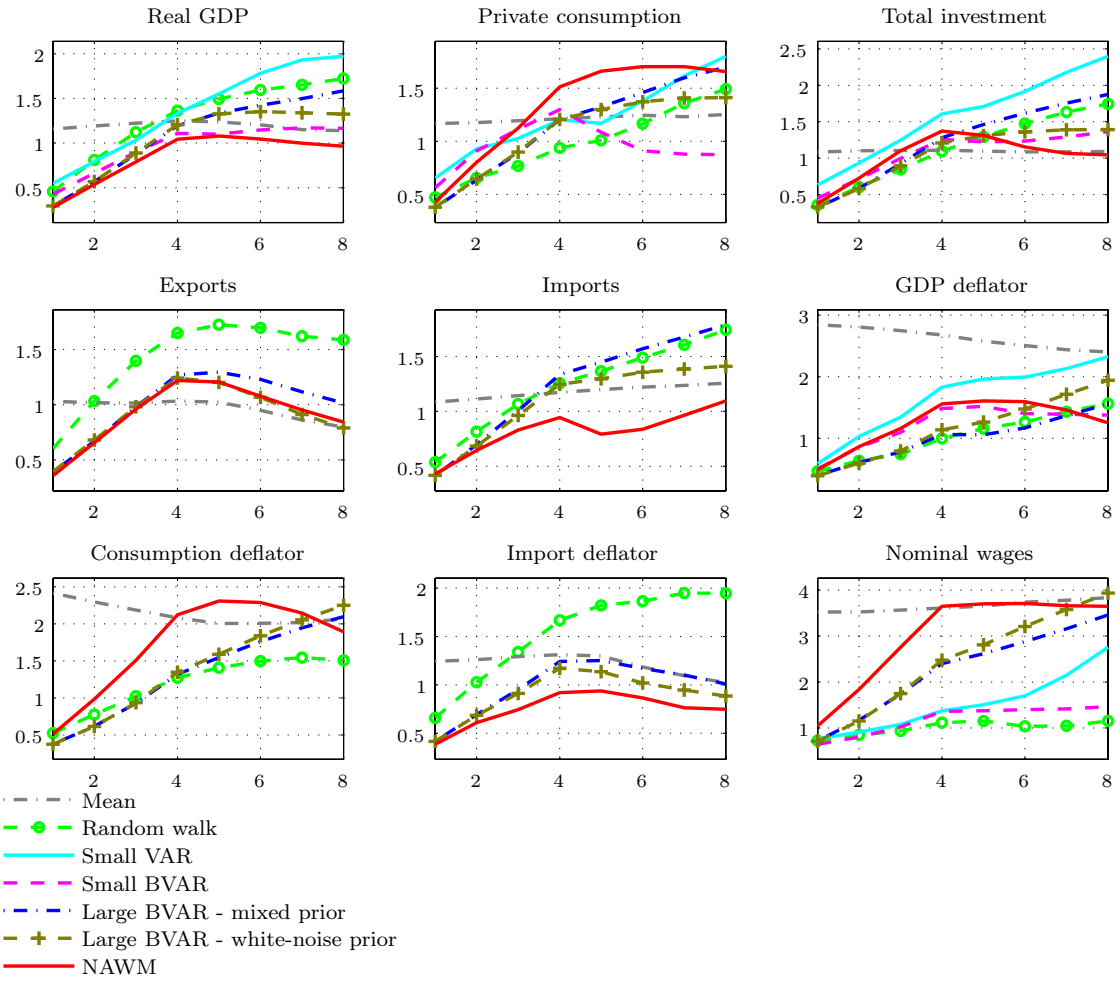


FIGURE 6. Trace statistics of the scaled MSE matrices when forecasting quarterly changes of variables in first differences.

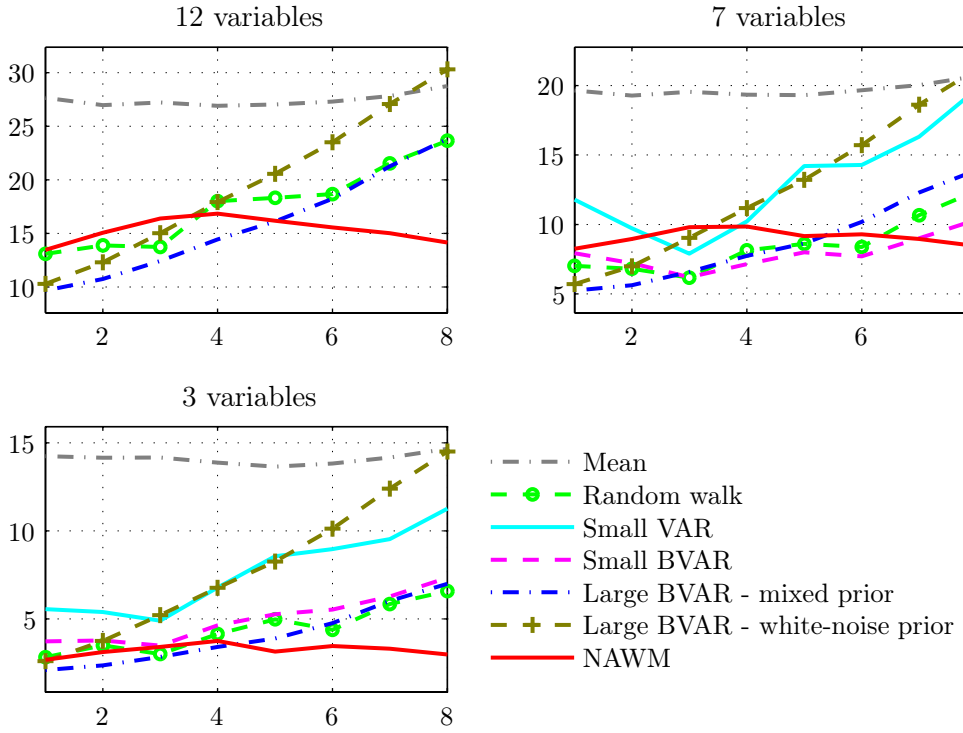


FIGURE 7. Trace statistics of the scaled MSE matrices when forecasting annual changes of variables in first differences.

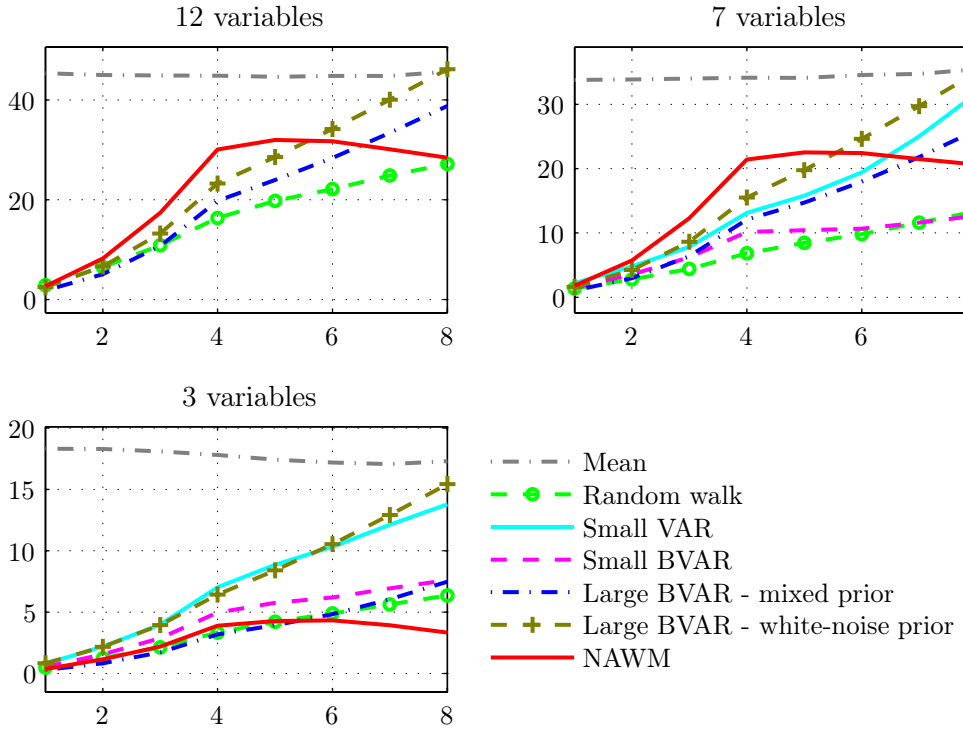


FIGURE 8. Log determinant statistics of the scaled MSE matrices when forecasting quarterly changes of variables in first differences.

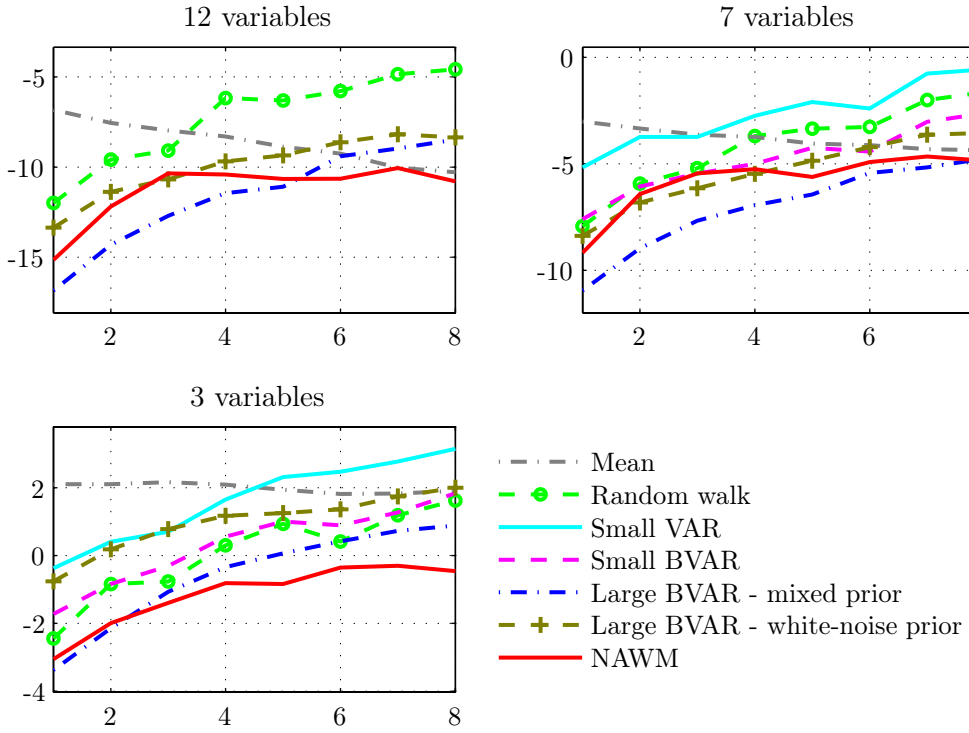


FIGURE 9. Log determinant statistics of the scaled MSE matrices when forecasting annual changes of variables in first differences.

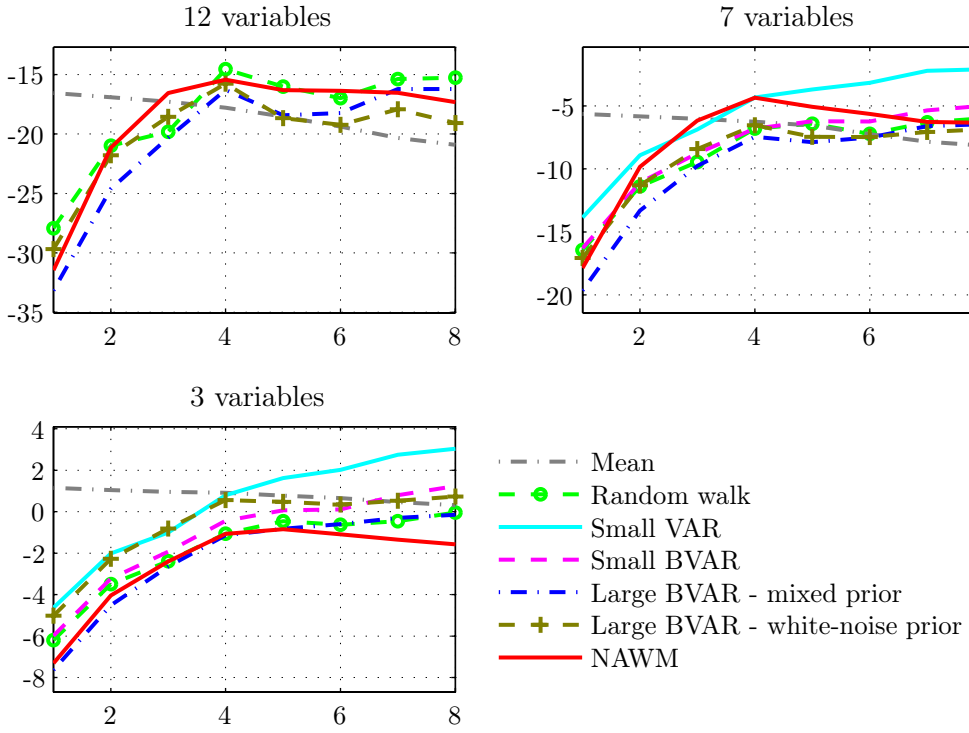


FIGURE 10. Log predictive scores when forecasting quarterly changes of variables in first differences.

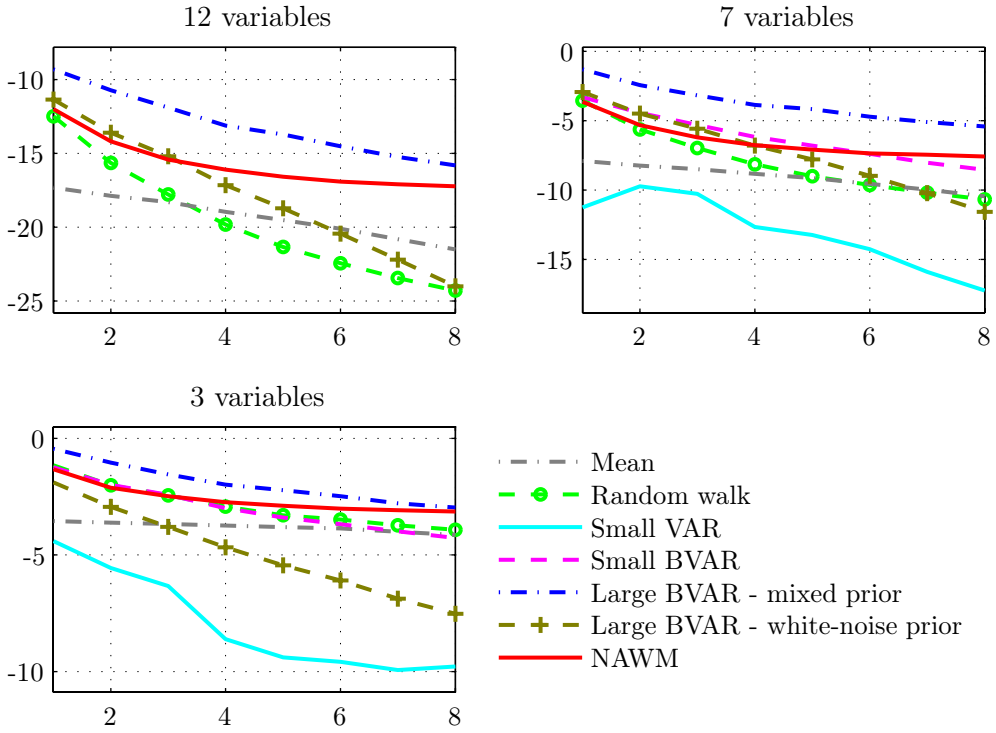


FIGURE 11. Log predictive scores when forecasting annual changes of variables in first differences.

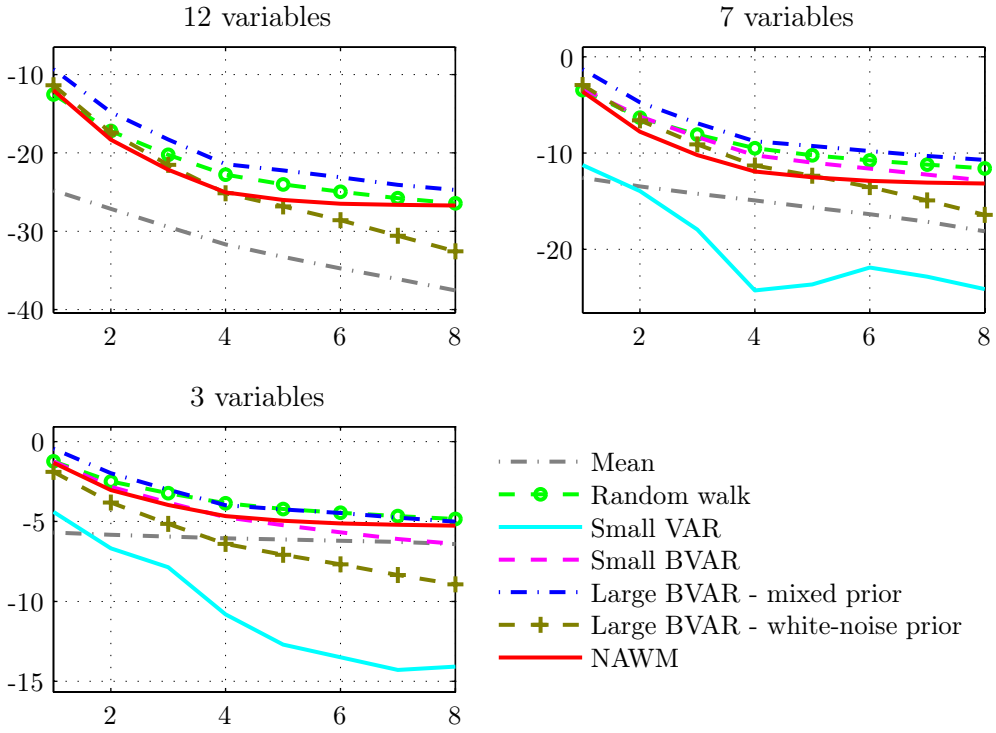
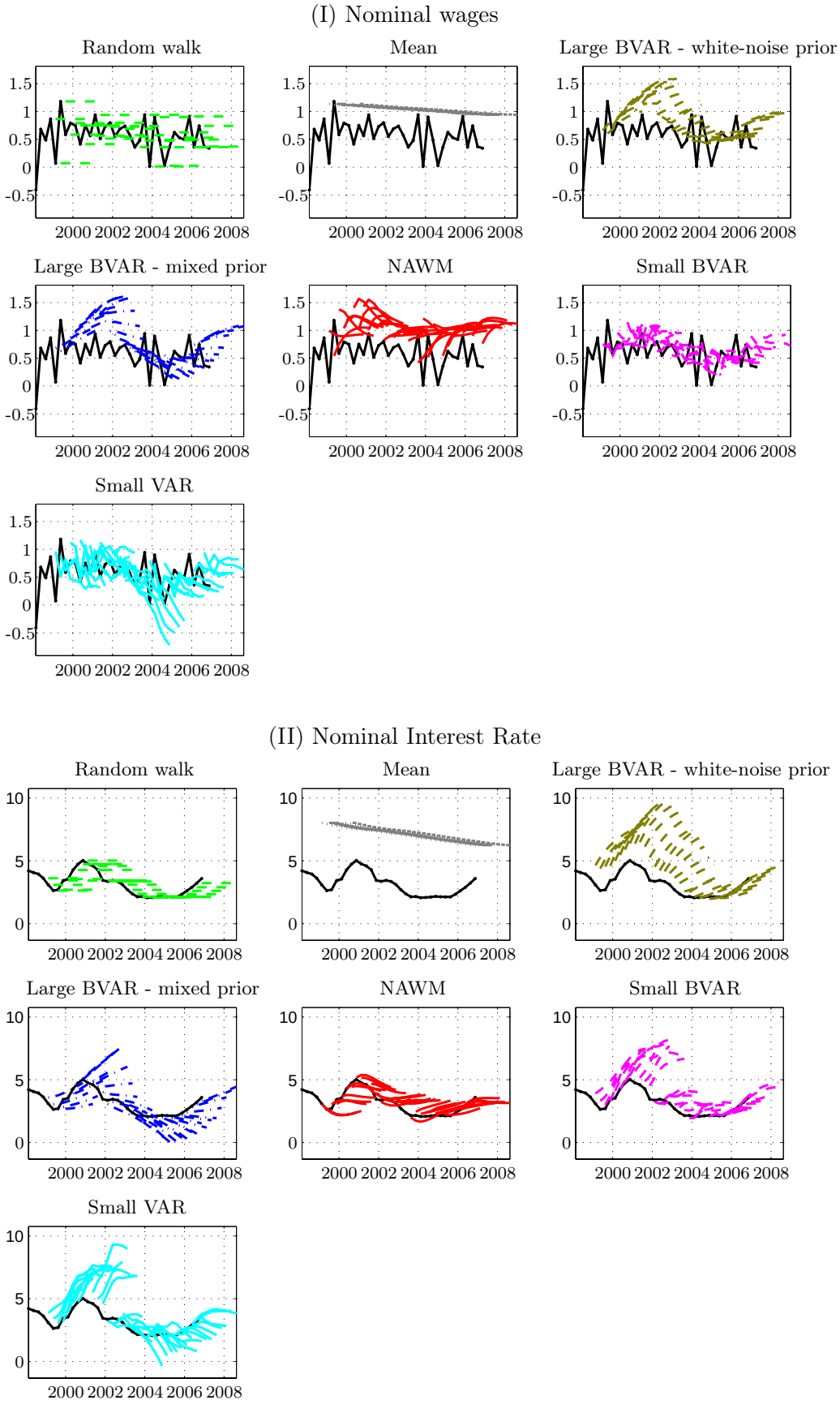


FIGURE 12. Quarterly nominal wage growth and nominal interest rate forecast paths.



References

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

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

Amisano, G., and R. Giacomini (2007): "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business & Economic Statistics*, 25, 177–190.

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

Anderson, G., and G. Moore (1985): "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models," *Economics Letters*, 17, 247–252.

Atkeson, A., and L. E. Ohanian (2001): "Are Phillps Curves Useful for Forecasting Inflation?," *Federal Reserve Bank of Minneapolis Quarterly Review*, 25, 2–11.

Bańbura, M., D. Giannone, and L. Reichlin (2008): "Large Bayesian VARs," ECB Working Paper Series No. 966.

Bartlett, M. S. (1957): "A Comment on D. V. Lindley's Statistical Paradox," *Biometrika*, 44, 533–534.

Bauwens, L., M. Lubrano, and J. F. Richard (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, Oxford.

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

Blanchard, O. J., and C. M. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305–1312.

Calvo, G. A. (1983): "Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, 12, 383–398.

Chib, S., and I. Jeliazkov (2001): "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.

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

Christiano, L. J., M. Trabandt, and K. Walentin (2009): "Introducing Financial Frictions and Unemployment into a Small Open Economy Model," Sveriges Riksbank Working Paper Series No. 214.

Christoffel, K., G. Coenen, and A. Warne (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.

Christoffersen, P. F. (1998): "Evaluating Interval Forecasts," *International Economic Review*, 39, 841–862.

Clements, M. P., and D. F. Hendry (1993): "On the Limitations of Comparing Mean Squared Forecast Errors," *Journal of Forecasting*, 12, 617–637, With discussion, p. 639–667.

——— (2005): "Evaluating a Model by Forecast Performance," *Oxford Bulletin of Economics and Statistics*, 67S, 931–956.

Clements, M. P., and J. Smith (2000): "Evaluating the Forecast Densities of Linear and Non-linear Models: Applications to Output Growth and Unemployment," *Journal of Forecasting*, 19, 255–276.

Dawid, A. P. (1984): "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society Series A*, 147, 278–292.

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

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

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

Diebold, F. (1998): "The Past, Present, and Future of Macroeconomic Forecasting," *Journal of Economic Perspectives*, 12(2), 175–192.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 337–374.

Diebold, F. X., A. S. Tay, and K. F. Wallis (1999): "Evaluating Density Forecasts of Inflation: The Survey of Professional Forecasters," in *Festschrift in Honour of C. W. J. Granger*, ed. by R. F. Engle, and H. White, pp. 76–90. Oxford University Press, Oxford, U.K.

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

Doan, T. A., R. B. Litterman, and C. A. Sims (1984): "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.

Edge, R. M., M. T. Kiley, and J.-P. Laforte (2009): "A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model," Federal Reserve Board Finance and Economics Discussion Series 2009-10.

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

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2005): "Estimating Dynamic Equilibrium Economies: Linear versus Non-Linear Likelihood," *Journal of Applied Econometrics*, 20, 891–910.

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

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

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

Geweke, J., and G. Amisano (2009): "Optimal Prediction Pools," ECB Working Paper Series No. 1017.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007): "Probabilistic Forecasts, Calibration and Sharpness," *Journal of the Royal Statistical Society Series B*, 69, 243–268.

Gneiting, T., and A. E. Raftery (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.

Gouriéroux, C., A. Monfort, and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S188.

Granger, C. W. J. (1999): *Empirical Modeling in Economics: Specification and Evaluation*. Cambridge University Press, Cambridge.

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7–46.

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

Kadiyala, K. R., and S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12, 99–132.

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.

Kydland, F. E., and E. C. Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–1370.

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

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

Lubik, T., and F. Schorfheide (2006): "A Bayesian Look at New Open Economy Macroeconomics," in *NBER Macroeconomics Annual 2005*, ed. by M. L. Gertler, and K. S. Rogoff, pp. 313–366. MIT Press.

Robertson, J. C., and E. W. Tallman (1999): "Vector Autoregressions: Forecasting and Reality," *Federal Reserve Bank of Atlanta Economic Review*, 84, 4–18.

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

Schorfheide, F. (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–670.

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

——— (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Sims, C. A., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949–968.

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

——— (2004): "Forecasting with a Bayesian DSGE Model: An Application to the Euro Area," *Journal of Common Market Studies*, 42, 841–867.

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

Smith, Jr., A. A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

Tay, A. S., and K. F. Wallis (2000): "Density Forecasting: A Survey," *Journal of Forecasting*, 19, 235–254.

Thompson, P. A., and R. B. Miller (1986): "Sampling the Future: A Bayesian Approach to Forecasting from Univariate Time Series Models," *Journal of Business & Economic Statistics*, 4, 427–436.

Villani, M. (2001): "Bayesian Prediction with Cointegrated Vector Autoregressions," *International Journal of Forecasting*, 17, 585–605.

——— (2009): "Steady-State Priors for Vector Autoregressions," *Journal of Applied Econometrics*, 24, 630–650.

Warne, A. (2009): "YADA Manual — Computational Details," Manuscript, European Central Bank. Available with the YADA distribution.