FORECASTING PERFORMANCE OF AN OPEN ECONOMY DYNAMIC STOCHASTIC GENERAL EQUILIBRIUM MODEL

MALIN ADOLFSON, JESPER LINDÉ, AND MATTIAS VILLANI

ABSTRACT. This paper analyzes the forecasting performance of an open economy DSGE model, estimated with Bayesian methods, for the Euro area during $1994Q1 - 2002Q4$. We compare the DSGE model and a few variants of this model to various reduced form forecasting models such as several vector autoregressions (VAR), estimated both by maximum likelihood and two different Bayesian approaches, and traditional benchmark models, e.g. the random walk. The accuracy of the point forecasts are assessed in a traditional out-of-sample rolling event evaluation using several univariate and multivariate measures. Forecast intervals are evaluated in different ways and the log predictive score is used to summarize the precision in the joint forecast distribution as a whole. We also discuss the role of Bayesian model probabilities and other frequently used low-dimensional summaries, e.g. the log determinant statistic, as measures of overall forecasting performance.

KEYWORDS: Bayesian inference, Forecasting; Open economy DSGE model; Vector autoregressive models.

JEL CLASSIFICATION: C11, C32, E37; E47.

1. INTRODUCTION

One of the objectives behind the formation of dynamic stochastic general equilibrium (DSGE) models is to explain and understand macroeconomic fluctuations using a coherent theoretical framework. The use of DSGE models in policy analysis, however, has been criticized by both academics and practitioners. The main argument has been the inability of DSGE models to - loosely speaking - fit the data. For instance, Pagan (2003) retains that there is a trade-off between theoretical and empirical coherence in DSGE models and VARs, the latter being more empirically than theoretically coherent relative to the former.

The new generation of DSGE models developed by Christiano, Eichenbaum and Evans (2005) among others, have shown great promise of improving the empirical properties by introducing nominal and real frictions into the model economy. Of course, the evaluation of fit can be assessed in various ways. For policy makers, the comparison of out-of-sample forecasting properties is of particular interest, as policy actions typically rely upon accurate assessments of the future development of the economy. Results in Smets and Wouters (2004) suggest that the new generation of closed economy DSGE models compare very well with vector autoregressive (VAR) models in terms of forecasting accuracy.

This paper evaluates the forecasting accuracy of an open economy DSGE model for the Euro area. This models enables us to predict several so called open economy variables such as, for example, the exchange rate, imports and exports. Evaluating the DSGE model for the latter variables are of particular interest, because previous research have demonstrated

Adolfson: *Sveriges Riksbank, malin.adolfson@riksbank.se;* Lindé: *Sveriges Riksbank and IIES, jesper.linde@riksbank.se;* Villani: *Sveriges Riksbank and Stockholm University*, *mattias.villani@riksbank.se.* The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank. Villani was partly financially supported by a grant from the Swedish Research Council (Vetenskapsrådet, grant no. 412-2002-1007).

the difficulties to project these variables accurately. By opening up the model economy we hope to better capture the workings of the real world economy, but it could very well be that the added complexity by itself deterioates the forecasting performance of the model. It is not uncommon to find that very small models are able to beat larger ones in forecasting. This motivates a thorough investigation of the model's forecasting performance with regards to both domestic and open economy macroeconomic variables.

A major difference between our analysis and Smets and Wouters' (2004), apart from the extension to the open economy setting, is that we include a unit-root stochastic technology shock, following Altig, Christiano, Eichenbaum and Lindé (2003). This induces a common stochastic trend in the variables and makes it possible to jointly model economic growth and business cycle fluctuations. In the empirical estimation and forecast evaluation we are hence not forced to detrend the data.

The DSGE model's forecasting properties are evaluated against a wide range of less theoretically oriented forecasting tools such as VARs, Bayesian VARs (BVARs), and naïve forecasts based on univariate random walks as well as on the simple means of the most recent data observations. Several authors have recently noted the theoretical connection between Bayesian model posterior probabilities and out-of-sample forecasting performance, e.g. Geweke (1999) and Del Negro, Schorfheide, Smets and Wouters (2004). Adding three alternative specifications of the benchmark DSGE model to the model set, allows us to study this link in some detail.

The forecasting performance of the models will be assessed in what is sometimes referred to as a rolling forecast evaluation. We use the observations in $1994Q1 - 2002Q4$ to evaluate the forecasts. We employ several univariate and multivariate measures to determine the accuracy of the point forecasts. Point forecasts are naturally the main concern of policy makers and has typically been the interest in the forecasting literature, see e.g. the M-competition in Makridakis et al. (1982). Recently, there has been a growing interest in forecast uncertainty. The so called fan charts used by Bank of England and Sveriges Riksbank (central bank of Sweden) to communicate the uncertainty in the inflation forecasts is one example. Using a Bayesian methodology we can derive the exact finite sample joint forecast distribution of all the endogenous variables in the system. We therefore also move beyond the evaluation of point forecasts to assess the reasonableness of, for example, predictive intervals.

The results indicate that the forecasting performance of the open economy DSGE model compares well with reduced form forecasting models such as VARs and BVARs. This holds true both in terms of the point forecast accuracy and when evaluating the accuracy of the whole forecast distribution. The paper also shows, using a spectral decomposition, that scalar valued multivariate measures of the forecasting performance should be interpreted with care since they run the risk of being dominated by dimensions in the set of projected variables which are minor interest to the policy maker.

The rest of the paper is organized as follows. Section 2 presents the theoretical DSGE model and reports the estimation results of four different specifications. In Section 3 we briefly discuss the alternative models used for forecasting such as vector autoregressive models and couple of naïve setups. Section 4 presents the accuracy measures that are subsequently employed in the empirical section. Section 5 reports the forecasting properties of the various theoretical and empirical models under consideration. Lastly, Section 6 summarizes and provides some conclusions.

2. The DSGE model

2.1. **Model.** This section gives an overview of the model economy and presents the key equations in the theoretical model. We refer to Adolfson, Laséen, Lindé and Villani (2005) for a more detailed description of the model.

The model economy includes four different categories of operating firms. These are domestic, importing consumption, importing investment, and exporting firms, respectively. Within each category there is a continuum of firms that each produces a differentiated good. The domestic firms produce their goods out of capital and labour inputs, and sell them to a retailer which transforms the intermediate products into a homogenous final good that is in turn sold to the households. Each importing firm (consumption and investment) buys a homogenous good in the world market and converts it into a differentiated good through a brand naming technology. An import consumption packer then aggregates the differentiated import consumption goods so that the final import consumption good is a composite of these differentiated products. Likewise, the imported investment goods are aggregated by an import investment packer. The exporting firms pursue a similar scheme. They buy the domestic final good, differentiate it and send their specific product to an export packer which aggregates the different export goods before the composite is sold to the consumers in the foreign market.

The final domestic good is a composite of a continuum of i differentiated goods, each supplied by a different firm, which follows the constant elasticity of substitution (CES) function

$$Y_t = \left[\int_0^1 (Y_{i,t})^{\frac{1}{\lambda_t^d}} \, di \right]^{\lambda_t^d}, 1 \leq \lambda_t^d < \infty, \tag{2.1}$$

where λ_t^d is a stochastic process that determines the time-varying markup in the domestic goods market. The demand for the differentiated product of the ith firm, $Y_{i,t}$, follows

$$Y_{i,t} = \left(\frac{P_{i,t}^d}{P_t^d} \right)^{-\frac{\lambda_t^d}{\lambda_t^d - 1}} Y_t. \tag{2.2}$$

The domestic production function for intermediate good i is given by

$$Y_{i,t} = z_t^{1-\alpha} \epsilon_t K_{i,t}^{\alpha} H_{i,t}^{1-\alpha} - z_t \phi, \tag{2.3}$$

where z_t is a unit-root technology shock, ϵ_t is a covariance stationary technology shock, and $H_{i,t}$ denotes homogeneous labour hired by the i^{th} firm. Notice that $K_{i,t}$ is not the physical capital stock, but rather the capital services stock, since we allow for variable capital utilization in the model. A fixed cost $z_t \phi$ is included in the production function. We set this parameter so that profits are zero in steady-state, following Christiano et al. (2005).

We allow for working capital by assuming that a fraction ν of the intermediate firms' wage bill has to be financed in advance. Cost minimization yields the following nominal marginal cost for intermediate firm i:

$$MC_t^d = \frac{1}{(1-\alpha)^{1-\alpha}} \frac{1}{\alpha^{\alpha}} (R_t^k)^{\alpha} \left[W_t (1 + \nu (R_{t-1} - 1)) \right]^{1-\alpha} \frac{1}{(z_t)^{1-\alpha}} \frac{1}{\epsilon_t}, \tag{2.4}$$

where R_t^k is the gross nominal rental rate per unit of capital services, R_{t-1} the gross nominal (economy wide) interest rate, and W_t the nominal wage rate per unit of aggregate, homogeneous, labour $H_{i,t}$.

Each of the domestic firms is subject to price stickiness through an indexation variant of the Calvo (1983) model. Since we have a time-varying inflation target in the model we allow for partial indexation to the current inflation target, but also to last period's inflation rate in order to allow for a lagged pricing term in the Phillips curve. Each intermediate firm faces in any period a random probability $(1-\xi_d)$ that it can reoptimize its price. The reoptimized price is denoted $P_t^{d,new}$.[1] The different firms maximize profits taking into account that there might not be a chance to optimally change the price in the future. Firm i therefore faces the following optimization problem when setting its price

$$\max_{P_t^{d,new}} \mathrm{E}_t \sum_{s=0}^{\infty} (\beta\xi_d)^s \upsilon_{t+s}[((\pi_t^d \pi_{t+1}^d ... \pi_{t+s-1}^d)^{\kappa_d} (\bar{\pi}_{t+1}^c \bar{\pi}_{t+2}^c ... \bar{\pi}_{t+s}^c)^{1-\kappa_d} P_t^{d,new})Y_{i,t+s} - MC_{i,t+s}^d (Y_{i,t+s} + z_{t+s}\phi^j)], \tag{2.5}$$

where the firm is using the stochastic discount factor $(\beta\xi_d)^s \upsilon_{t+s}$ to make profits conditional upon utility. β is the discount factor, and υ_{t+s} the marginal utility of the households' nominal income in period $t+s$, which is exogenous to the intermediate firms. π_t^d denotes inflation in the domestic sector, $\bar{\pi}_t^c$ a time-varying inflation target of the central bank and $MC_{i,t}^d$ the nominal marginal cost.

The first order condition of the profit maxmization problem in equation (2.5) yields the following log-linearized Phillips curve:

$$\begin{aligned} \left(\widehat{\pi}_t^d - \widehat{\bar{\pi}}_t^c\right) = {} & \frac{\beta}{1+\kappa_d\beta}\left(\mathrm{E}_t\widehat{\pi}_{t+1}^d - \rho_\pi \widehat{\bar{\pi}}_t^c\right) + \frac{\kappa_d}{1+\kappa_d\beta}\left(\widehat{\pi}_{t-1}^d - \widehat{\bar{\pi}}_t^c\right) \\ & - \frac{\kappa_d\beta(1-\rho_\pi)}{1+\kappa_d\beta}\widehat{\bar{\pi}}_t^c + \frac{(1-\xi_d)(1-\beta\xi_d)}{\xi_d(1+\kappa_d\beta)}\left(\widehat{mc}_t^d + \widehat{\lambda}_t^d\right), \end{aligned} \tag{2.6}$$

where a hat denotes log-linearized variables (i.e., $\hat{X}_t = dX_t/X$).

We now turn to the import and export sectors. There is a continuum of importing consumption and investment firms that buy a homogenous good at price P_t^* in the world market, and differentiate this good by brand naming. The exporting firms buy the (homogenous) domestic final good at price P_t^d and turn this into a differentiated export good through the same type of brand naming technology. The nominal marginal cost of the importing and exporting firms are thus $S_t P_t^*$ and P_t^d/S_t, respectivley. The differentiated import and export goods are subsequently aggregated by an import consumption, import investment and export packer, respectively, so that the final import consumption, import investment, and export good is each a CES composite according to the following:

$$C_t^m = \left[\int_0^1 (C_{i,t}^m)^{\frac{1}{\lambda_t^{mc}}} di\right]^{\lambda_t^{mc}}, \quad I_t^m = \left[\int_0^1 (I_{i,t}^m)^{\frac{1}{\lambda_t^{mi}}} di\right]^{\lambda_t^{mi}}, \quad X_t = \left[\int_0^1 (X_{i,t})^{\frac{1}{\lambda_t^{x}}} di\right]^{\lambda_t^{x}}, \tag{2.7}$$

where $1 \leq \lambda_t^j < \infty$ for $j = \{mc, mi, x\}$ is the time-varying markup in the import consumption (mc), import investment (mi) and export (x) sector. By assumption the continuum of consumption and investment importers invoice in the domestic currency and exporters in the foreign currency. In order to allow for short-run incomplete exchange rate pass-through to import as well as export prices we therefore introduce nominal rigidities in the local currency price. This is modeled through the same type of Calvo setup as above. The price setting

[1] For the firms that are not allowed to reoptimize their price, we adopt the indexation scheme $P_{t+1}^d = (\pi_t^d)^{\kappa_d} (\bar{\pi}_{t+1}^c)^{1-\kappa_d} P_t^d$ where κ_d is an indexation parameter.

problems of the importing and exporting firms are completely analogous to that of the domestic firms in equation (2.5), and the demand for the differentiated import and export goods follow similar expressions as to equation (2.2). In total there is thus four specific Phillips curve relations determining inflation in the domestic, import consumption, import investment and export sectors.

In the model economy there is also a continuum of households which attain utility from consumption, leisure and real cash balances. The preferences of household j are given by

$$\mathrm{E}_0^j \sum_{t=0}^{\infty} \beta^t \left[\zeta_t^c U\left(C_{j,t} - bC_{j,t-1}\right) - \zeta_t^h A_L \frac{\left(h_{j,t}\right)^{1+\sigma_L}}{1+\sigma_L} + A_q \frac{\left(\frac{Q_{j,t}}{z_t P_t^d}\right)^{1-\sigma_q}}{1-\sigma_q} \right], \tag{2.8}$$

where $C_{j,t}$, $h_{j,t}$ and $Q_{j,t}/P_t^d$ denote the j^{th} household's levels of aggregate consumption, labour supply and real cash holdings, respectively. Consumption is subject to habit formation through $bC_{j,t-1}$. ζ_t^c and ζ_t^h are persistent preference shocks to consumption and labour supply, respectively. To make cash balances in equation (2.8) stationary when the economy is growing they are scaled by the unit root technology shock z_t. Households consume a basket of domestically produced goods and imported products which are supplied by the domestic and importing consumption firms, respectively. Aggregate consumption is assumed to be given by the following constant elasticity of substitution (CES) function:

$$C_t = \left[(1-\omega_c)^{1/\eta_c} \left(C_t^d\right)^{(\eta_c-1)/\eta_c} + \omega_c^{1/\eta_c} \left(C_t^m\right)^{(\eta_c-1)/\eta_c} \right]^{\eta_c/(\eta_c-1)}, \tag{2.9}$$

where C_t^d and C_t^m are consumption of the domestic and imported good, respectively. ω_c is the share of imports in consumption, and η_c is the elasticity of substitution across consumption goods.

The households invest in a basket of domestic and imported investment goods to form the physical capital stock, and decide how much capital services to rent to the domestic firms, given certain capital adjustment costs. These are costs to adjusting the investment rate as well as costs of varying the utilization rate of the physical capital stock. The households can increase their capital stock by investing in additional physical capital (I_t), taking one period to come in action, or by directly increasing the utilization rate of the capital at hand ($u_t = K_t/\bar{K}_t$). The capital accumulation equation for the physical capital stock ($\bar{K}_t$) is given by

$$\bar{K}_{t+1} = (1-\delta)\bar{K}_t + \Upsilon_t \left(1 - \tilde{S}\left(I_t/I_{t-1}\right)\right) I_t, \tag{2.10}$$

where $\tilde{S}\left(I_t/I_{t-1}\right)$ determines the investment adjustment costs through the estimated parameter $\tilde{S}''$, and Υ_t is a stationary investment-specific technology shock. Total investment is assumed to be given by a CES aggregate of domestic and imported investment goods (I_t^d and I_t^m, respectively) according to

$$I_t = \left[(1-\omega_i)^{1/\eta_i} \left(I_t^d\right)^{(\eta_i-1)/\eta_i} + \omega_i^{1/\eta_i} \left(I_t^m\right)^{(\eta_i-1)/\eta_i} \right]^{\eta_i/(\eta_i-1)}, \tag{2.11}$$

where ω_i is the share of imports in investment, and η_i is the elasticity of substitution across investment goods.

Further, along the lines of Erceg, Henderson and Levin (2000), each household is a monopoly supplier of a differentiated labour service which implies that they can set their own wage. After having set their wage, households inelastically supply the firms' demand for labour at the going wage rate. Each household sells its labour to a labour packing firm which transforms household labour into a homogenous good that is demanded by each of the domestic goods producing firms. Wage stickiness is introduced through the Calvo (1983) setup, with partial indexation to last period's CPI inflation rate, the current inflation target and the technology growth. Household j reoptimizes its nominal wage rate $W_{j,t}^{new}$ according to the following

$$\begin{aligned} \max_{W_{j,t}^{new}} \quad & \mathrm{E}_t \sum_{s=0}^{\infty} (\beta\xi_w)^s \left[-\zeta_{t+s}^h A_L \frac{(h_{j,t+s})^{1+\sigma_L}}{1+\sigma_L} + \right. \\ & \left. \upsilon_{t+s} \frac{(1-\tau_{t+s}^y)}{(1+\tau_{t+s}^w)} \left(\left(\pi_t^c ... \pi_{t+s-1}^c\right)^{\kappa_w} \left(\bar{\pi}_{t+1}^c ... \bar{\pi}_{t+s}^c\right)^{(1-\kappa_w)} \left(\mu_{z,t+1} ... \mu_{z,t+s}\right) W_{j,t}^{new} \right) h_{j,t+s} \right], \end{aligned} \tag{2.12}$$

where ξ_w is the probability in any period that a household is not allowed to reoptimize its wage, τ_t^y a labour income tax, τ_t^w a pay-roll tax (paid for simplicity by the households), and $\mu_{z,t} = z_t/z_{t-1}$ the growth rate of the permanent technology level.[2]

The households can save in domestic bonds and foreign bonds, and also hold cash. This choice balances into an arbitrage condition pinning down expected exchange rate changes (i.e., an uncovered interest rate parity condition). To ensure a well-defined steady-state in the model, we allow for imperfect financial integration in the international fincancial markets. We assume that there is a premium on the foreign bond holdings which depends on the aggregate net foreign asset position of the domestic households, following Lundvik (1992) and Benigno (2001):

$$\Phi(a_t, \tilde{\phi}_t) = \exp(-\tilde{\phi}_a(a_t - \bar{a}) + \tilde{\phi}_t), \tag{2.13}$$

where $A_t \equiv (S_t B_t^*)/(P_t z_t)$ is the net foreign asset position, and $\tilde{\phi}_t$ a shock to the risk premium.

The budget constraint for the households given by

$$\begin{aligned} & M_{t+1} + S_t B_{t+1}^* + P_t^c C_t (1+\tau_t^c) + P_t^i I_t + P_t a(u_t) \overline{K}_t \\ = \; & R_{t-1}(M_t - Q_t) + Q_t + R_{t-1}^* \Phi(a_{t-1}, \widetilde{\phi}_{t-1}) S_t B_t^* \\ & + \left(1-\tau_t^k\right) R_t^k u_t \bar{K}_t + (1-\tau_t^y) \frac{W_t}{1+\tau_t^w} h_t + \left(1-\tau_t^k\right) \Pi_t \\ & -\tau_t^k \left[(R_{t-1}-1)(M_t - Q_t) + \left(R_{t-1}^* \Phi(a_{t-1}, \widetilde{\phi}_{t-1}) - 1\right) S_t B_t^* + B_t^* (S_t - S_{t-1}) \right] \\ & + TR_t + D_t, \end{aligned} \tag{2.14}$$

where the right-hand side describes the resources at disposal. The households earn interest on the amount of nominal domestic assets that are not held as cash, $M_t - Q_t$. They can also save in foreign bonds B_t^*, which pay a risk-adjusted pre-tax gross interest rate of $R_{t-1}^* \Phi(a_{t-1}, \widetilde{\phi}_{t-1})$. S_t is the nominal exchange rate (foreign currency per unit of domestic currency), R_t^k the gross rental rate of capital, and W_t the nominal wage rate. The households earn income from renting capital and labour services (K_t and h_t) to the intermediate firms, where u_t denotes the varying capital utilization rate and $\bar{K}_t$ the physical capital stock. They pay taxes on consumption (τ_t^c), capital income (τ_t^k), labour income (τ_t^y), and on the pay-roll (τ_t^w). Π_t denotes profits, TR_t lump-sum transfers from the government, and D_t the household's net cash income from participating in state contingent securities at time t. The right hand side describes how the households spend their resources on consumption and investment goods, priced at P_t^c and P_t^i respectively, on future bond holdings, and pay the cost of varying the capital utilization rate $P_t a(u_t)\bar{K}_t$, where $a(u_t)$ is the utilization cost function.

[2]For the households that are not allowed to reoptimize, the indexation scheme is $W_{j,t+1} = (\pi_t^c)^{\kappa_w} (\bar{\pi}_{t+1}^c)^{(1-\kappa_w)} \mu_{z,t+1} W_{j,t}^{new}$, where κ_w is an indexation parameter.

Following Smets and Wouters (2003), monetary policy is approximated with the instrument rule (expressed in log-linearized terms)

$$\begin{aligned}(2.15)\qquad \widehat{R}_t &= \rho_R \widehat{R}_{t-1} + (1-\rho_R)\left[\widehat{\bar{\pi}}^c_t + r_\pi\left(\hat{\pi}^c_{t-1} - \widehat{\bar{\pi}}^c_t\right) + r_y \hat{y}_{t-1} + r_x \hat{x}_{t-1}\right] \\ &\quad + r_{\Delta\pi}\left(\hat{\pi}^c_t - \hat{\pi}^c_{t-1}\right) + r_{\Delta y}\Delta \hat{y}_t + \varepsilon_{R,t},\end{aligned}$$

where $\varepsilon_{R,t}$ is an uncorrelated monetary policy shock. Thus, the central bank is assumed to adjust the short term interest rate in response to deviations of CPI inflation from the time-varying inflation target $\left(\hat{\pi}^c_t - \widehat{\bar{\pi}}^c_t\right)$, the output gap ($\hat{y}_t$, measured as actual minus trend output), the real exchange rate ($\hat{x}_t$) and the interest rate set in the previous period. In addition, note that the nominal interest rate adjusts directly to the inflation target.

To clear the final goods market, the foreign bond market, and the loan market, the following three constraints must hold in equilibrium:

$$(2.16)\qquad C^d_t + I^d_t + G_t + C^x_t + I^x_t \leq z_t^{1-\alpha}\epsilon_t K^\alpha_t H^{1-\alpha}_t - z_t\phi - a(u_t)\bar{K}_t,$$

$$(2.17)\qquad S_t B^*_{t+1} = S_t P^x_t\left(C^x_t + I^x_t\right) - S_t P^*_t\left(C^m_t + I^m_t\right) + R^*_{t-1}\Phi(a_{t-1}, \widetilde{\phi}_{t-1}) S_t B^*_t,$$

$$(2.18)\qquad \nu W_t H_t = \mu_t M_t - Q_t,$$

where C^x_t and I^x_t are the foreign demand for export goods, P^*_t the foreign price level, and $\mu_t = M_{t+1}/M_t$ is the monetary injection by the central bank. When defining the demand for export goods, we introduce a stationary asymmetric technology shock $\tilde{z}^*_t = z^*_t/z_t$, where z^*_t is the permanent technology level abroad, to allow for different degrees of technological progress domestically and abroad.

The structural shock processes in the model is given in log-linearized form by the univariate representation

$$\hat{x}_t = \rho_x \hat{x}_{t-1} + \varepsilon_{x,t}, \qquad \varepsilon_{x,t} \overset{iid}{\sim} N\left(0, \sigma^2_x\right)$$

where $x = \{\ \mu_{z,t},\ \epsilon_t, \lambda^j_t,\ \zeta^c_t,\ \zeta^h_t,\ \Upsilon_t,\ \tilde{\phi}_t,\ \varepsilon_{R,t}, \bar{\pi}^c_t,\ \tilde{z}^*_t\}$ and $j = \{d, mc, mi, x\}$.

Lastly, to simplify the analysis we adopt the assumption that the foreign prices, output (HP-detrended) and interest rate are exogenously given by an identified VAR(4) model. The fiscal policy variables - taxes on capital income, labour income, consumption, and the pay-roll, together with (HP-detrended) government expenditures - are assumed to follow an identified VAR(2) model.[3]

2.2. **Calibration and estimation.** In order to efficiently compute the likelihood function, the model is log-linearized and the reduced form of the model is obtained by the AIM algorithm developed by Anderson and Moore (1985). As a first step, we cast the log-linearized model on matrix form as

$$(2.19)\qquad \mathrm{E}_t\left\{\alpha_0 \tilde{z}_{t+1} + \alpha_1 \tilde{z}_t + \alpha_2 \tilde{z}_{t-1} + \beta_0 \theta_{t+1} + \beta_1 \theta_t\right\} = 0,$$

where $\tilde{z}_t$ is a $n_{\tilde{z}} \times 1$ vector with log-linearized endogenous variables and θ_t is a $n_\theta \times 1$ vector with exogenous variables which follows

$$(2.20)\qquad \theta_t = \rho\theta_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \Sigma).$$

Note that since some of the processes for the exogenous variables are given by more than one lag, we expand θ_t with lags of the relevant exogenous variable.

[3]It should be noted that Adolfson et al. (2005) report that the fiscal shocks have small dynamic effects in the model. This is because households are Ricardian and infinitively lived, and these shocks are transitory with relatively small variance in the data, which together do not generate any wealth effects.

The solution of the fundamental difference equation can then be written as

$$\tilde{z}_t = A\tilde{z}_{t-1} + B\theta_t \tag{2.21}$$

where A and B are the so called feedback and feed-forward matrices, respectively.

The solution of the model given by (2.21) and (2.20) can be transformed to the following state-space representation for the partially unobserved state variables $\xi_t = (z_t,\ \theta_t)'$ in the model

$$\xi_t = F_\xi \xi_{t-1} + v_t, \tag{2.22}$$

where $v_t \overset{iid}{\sim} N(0, Q)$, and the observation equation can be written

$$\tilde{Y}_t = A'_X X_t + H'\xi_t + \zeta_t, \tag{2.23}$$

where $\tilde{Y}_t$ is a vector of observed variables, X_t a vector with exogenous variables (e.g., a constant) and $\zeta_t \overset{iid}{\sim} N(0, R)$.

In order to facilitate identification of the various shocks and parameters that we estimate (we estimate 11 shocks that follow AR(1) processes, and 2 shocks that are assumed to be i.i.d.), we include the following set of 15 observable variables in $\tilde{Y}_t$ in (2.23): the domestic inflation rate, the short-run interest rate, employment, consumption, investment, GDP, the real wage, exports, imports, the consumption deflator and the investment deflator, the real exchange rate, foreign inflation, the foreign interest rate, and foreign output.[4] Despite the fact that the foreign variables are exogenous, we still include them as observable variables as they enable identification of the asymmetric technology shock and are informative about the parameters governing the transmission of foreign impulses to the domestic economy.

To make the data stationary we experiment with two different strategies. In the first strategy, all real variables enter $\tilde{Y}_t$ in first differences. It is important to note that the unit root technology shock in the theoretical model induces a common stochastic trend in the levels of the real variables. In the second strategy, we therefore exploit the cointegration structure of the theoretical model and all real variables except GDP enter $\tilde{Y}_t$ as deviations from the GDP level, while GDP itself enters in first difference form. In Figure 1 the data series are depicted with real variables in yearly growth rates. Note that employment and the real exchange rate are measured as percentage deviations around the mean.

[4]The data set employed here was first constructed by Fagan et al. (2001). The Fagan data set includes foreign (i.e., rest of the world) output and inflation, but not a foreign interest rate. We therefore use the Fed funds rate as a proxy for this series. Note also that there is no (official) data on aggregate hours worked, $\hat{H}_t$, available for the euro area. Therefore, we use employment in our estimations. Since employment is likely to respond more slowly to shocks than hours worked, we model employment using Calvo-rigidity (following Smets and Wouters, 2003). For reasons discussed in greater detail in Adolfson et al. (2005), we take out a linear trend in employment and the excess trend in imports and exports relative to the trend in GDP prior to estimation.

Table 1: Calibrated parameters

Parameter	Description	Calibrated value
β	Households' discount factor	0.999
α	Capital share of income	0.29
η_c	Substitution elasticity between C_t^d and C_t^m	5.00
σ_a	Capital utilization cost parameter	10^6
μ	Money growth rate (quarterly rate)	1.01
σ_L	Labor supply elasticity	1.00
δ	Depreciation rate	0.013
λ_w	Wage markup	1.05
ω_i	Share of imported investment goods	0.55
ω_c	Share of imported consumption goods	0.31
ν	Share of wage bill financed by loans	1.00
τ^y	Labor income tax rate	0.177
τ^c	Value added tax rate	0.125
$\rho_{\bar{\pi}}$	Inflation target persistence	0.975
g_r	Government expenditures-output ratio	0.204
	Implied steady state relationships*	
$\bar{\pi}$	Steady state inflation rate (percent)	2.02
R	Nominal interest rate (percent)	5.30
C/Y	Consumption-output ratio	0.58
I/Y	Investment-output ratio	0.22
$\tilde{X}/Y = \tilde{M}/Y$	Export/Import output ratio	0.25
$S_t = S_{t+1}$	Nominal exchange rate	1.00
A	Net foreign assets	0.00
X	Real exchange rate	1.00

Note: The steady state is affected by some parameters that are estimated, e.g. μ_z, λ_d, $\lambda_{m,c}$ and $\lambda_{m,i}$, which implies that the steady state values differ somewhat between the prior and the posterior. The table reports the implied values given by these parameters evaluated at the prior mode.

A subset of the parameters are calibrated (infinitely strict priors), whereas another subset of parameters are estimated using Bayesian techniques. We choose to calibrate those parameters which we think are weakly identified by the variables that we include in $\tilde{Y}_t$. These parameters are mostly related to the steady-state values (i.e., the great ratios). Table 1 reports the calibrated parameters along with the implied steady state values of some key variables. The remaining parameters are estimated. The estimated parameters pertain mostly to the nominal and real frictions in the model as well as the exogenous shock processes described above.

Table 2 shows the assumptions for the prior distribution of the estimated parameters. The location of the prior distribution of the 51 estimated parameters corresponds to a large extent to those in Smets and Wouters (2003) and the findings in Altig et al. (2003) on U.S. data. See Adolfson et al. (2005) for a more detailed discussion about our choice of prior distributions.

The joint posterior distribution of all estimated parameters is obtained in two steps. First, the posterior mode and Hessian matrix evaluated at the mode is computed by standard numerical optimization routines. We use the first 10 years of the full sample $1970Q1 - 2002Q4$ to obtain a prior on the unobserved state, and use the subsample $1980Q1 - 2002Q4$ for inference.

Table 2: Prior and posterior distributions

Parameter		Prior distribution			Posterior distribution					
					First differences					Levels with cointegration
					No variable capital utilization		Variable capital utiliz.	Persistent markup shock	IID markup shocks	
					$\sigma_a = 10^6$		$\sigma_a =$ 0.049	$\rho_{\lambda_d} > 0$	$\rho_{\lambda_d} = \rho_{\lambda_{mc}} = \rho_{\lambda_{mi}} = \rho_{\lambda_x} = 0$	$\sigma_a = 10^6$
		type	mean	std.dev. /df	mode	std. dev. (Hessian)	mode	mode	mode	mode
Calvo wages	ξ_w	beta	0.675	0.050	0.697	0.047	0.716	0.626	0.687	0.693
Calvo domestic prices	ξ_d	beta	0.675	0.050	0.883	0.015	0.895	0.661	0.882	0.930
Calvo import cons. prices	$\xi_{m,c}$	beta	0.500	0.100	0.463	0.059	0.523	0.523	0.899*	0.619
Calvo import inv. prices	$\xi_{m,i}$	beta	0.500	0.100	0.740	0.040	0.743	0.714	0.912*	0.774
Calvo export prices	ξ_x	beta	0.500	0.100	0.639	0.059	0.630	0.669	0.853*	0.649
Calvo employment	ξ_e	beta	0.675	0.100	0.792	0.022	0.757	0.795	0.784	0.833
Indexation wages	κ_w	beta	0.500	0.150	0.516	0.160	0.453	0.291	0.480	0.592
Indexation domestic prices	κ_d	beta	0.500	0.150	0.212	0.066	0.173	0.171	0.188	0.200
Index. import cons. prices	$\kappa_{m,c}$	beta	0.500	0.150	0.161	0.074	0.128	0.148	0.256	0.145
Index. import inv. prices	$\kappa_{m,i}$	beta	0.500	0.150	0.187	0.079	0.192	0.200	0.830	0.202
Indexation export prices	κ_x	beta	0.500	0.150	0.139	0.072	0.148	0.125	0.262	0.148
Markup domestic	λ_d	inv. gamma	1.200	2	1.168	0.053	1.174	1.155	1.160	1.169
Markup imported cons.	$\lambda_{m,c}$	inv. gamma	1.200	2	1.619	0.063	1.636	1.642	1.515	1.535
Markup.imported invest.	$\lambda_{m,i}$	inv. gamma	1.200	2	1.226	0.088	1.209	1.255	1.160	1.164
Investment adj. cost	$\tilde{S}''$	normal	7.694	1.500	8.732	1.370	9.052	7.143	9.499	8.227
Habit formation	b	beta	0.650	0.100	0.690	0.048	0.694	0.614	0.647	0.761
Subst. elasticity invest.	η_i	inv. gamma	1.500	4	1.669	0.273	1.585	1.616	1.405	1.526
Subst. elasticity foreign	η_f	inv. gamma	1.500	4	1.460	0.098	1.400	1.577	1.356	1.416
Technology growth	μ_z	trunc. normal	1.006	0.0005	1.005	0.000	1.005	1.006	1.005	1.006
Capital income tax	τ_k	beta	0.120	0.050	0.137	0.042	0.220	0.265	0.172	0.241
Labour pay-roll tax	τ_w	beta	0.200	0.050	0.186	0.050	0.183	0.185	0.186	0.185
Risk premium	$\tilde{\phi}$	inv. gamma	0.010	2	0.145	0.047	0.131	0.095	0.035	0.095
Unit root tech. shock	ρ_{μ_z}	beta	0.850	0.100	0.723	0.106	0.753	0.792	0.741	0.810
Stationary tech. shock	ρ_ε	beta	0.850	0.100	0.909	0.030	0.935	0.997	0.904	0.980
Invest. spec. tech shock	ρ_Υ	beta	0.850	0.100	0.750	0.041	0.738	0.562	0.785	0.799
Asymmetric tech. shock	$\rho_{\tilde{z}*}$	beta	0.850	0.100	0.993	0.002	0.992	0.953	0.990	0.993
Consumption pref. shock	ρ_{ζ_c}	beta	0.850	0.100	0.935	0.029	0.935	0.992	0.911	0.990
Labour supply shock	ρ_{ζ_h}	beta	0.850	0.100	0.675	0.062	0.646	0.536	0.656	0.824
Risk premium shock	$\rho_{\tilde{\phi}}$	beta	0.850	0.100	0.991	0.008	0.990	0.991	0.920	0.995
Domestic markup shock	ρ_{λ_d}							0.995		
Imp. cons. markup shock	$\rho_{\lambda_{m,c}}$	beta	0.850	0.100	0.978	0.016	0.984	0.975		0.960
Imp. invest. markup shock	$\rho_{\lambda_{m,i}}$	beta	0.850	0.100	0.974	0.015	0.971	0.990		0.983
Export markup shock	ρ_{λ_x}	beta	0.850	0.100	0.894	0.045	0.895	0.928		0.906
Unit root tech. shock	σ_z	inv. gamma	0.200	2	0.130	0.025	0.122	0.132	0.128	0.120
Stationary tech. shock	σ_ε	inv. gamma	0.700	2	0.452	0.082	0.414	0.422	0.450	0.467
Invest. spec. tech. shock	σ_Υ	inv. gamma	0.200	2	0.424	0.046	0.397	0.444	0.376	0.328
Asymmetric tech. shock	$\sigma_{\tilde{z}*}$	inv. gamma	0.400	2	0.203	0.031	0.200	0.186	0.204	0.343
Consumption pref. shock	σ_{ζ_c}	inv. gamma	0.200	2	0.151	0.031	0.132	0.155	0.163	0.113
Labour supply shock	σ_{ζ_h}	inv. gamma	0.200	2	0.095	0.015	0.094	0.098	0.096	0.085
Risk premium shock	$\sigma_{\tilde{\phi}}$	inv. gamma	0.050	2	0.130	0.023	0.123	0.122	0.344	0.119
Domestic markup shock	σ_λ	inv. gamma	0.300	2	0.130	0.012	0.133	0.125	0.129	0.122
Imp. cons. markup shock	$\sigma_{\lambda_{m,c}}$	inv. gamma	0.300	2	2.548	0.710	1.912	1.810	1.147	0.943
Imp. invest. markup shock	$\sigma_{\lambda_{m,i}}$	inv. gamma	0.300	2	0.292	0.079	0.281	0.341	0.414	0.219
Export markup shock	σ_{λ_x}	inv. gamma	0.300	2	0.977	0.214	1.028	0.789	1.272	0.893
Monetary policy shock	σ_R	inv. gamma	0.150	2	0.133	0.013	0.126	0.144	0.130	0.125
Inflation target shock	$\sigma_{\bar{\pi}^c}$	inv. gamma	0.050	2	0.044	0.012	0.036	0.041	0.049	0.056
Interest rate smoothing	ρ_R	beta	0.800	0.050	0.874	0.021	0.885	0.824	0.851	0.834
Inflation response	r_π	normal	1.700	0.100	1.710	0.067	1.615	1.660	1.697	1.687
Diff. infl response	$r_{\Delta\pi}$	normal	0.300	0.100	0.317	0.059	0.301	0.384	0.304	0.388
Real exch. rate response	r_x	normal	0.000	0.050	-0.009	0.008	-0.010	-0.008	0.003	-0.012
Output response	r_y	normal	0.125	0.050	0.078	0.028	0.123	-0.030	0.056	0.059
Diff. output response	$r_{\Delta\pi}$	normal	0.0625	0.050	0.116	0.028	0.142	0.130	0.104	0.129
Log marginal likelihood					-1909.34		-1917.39	-1915.53	-1975.5	-1953.70

*Note: The same prior is used as for the domestic price stickiness parameter.

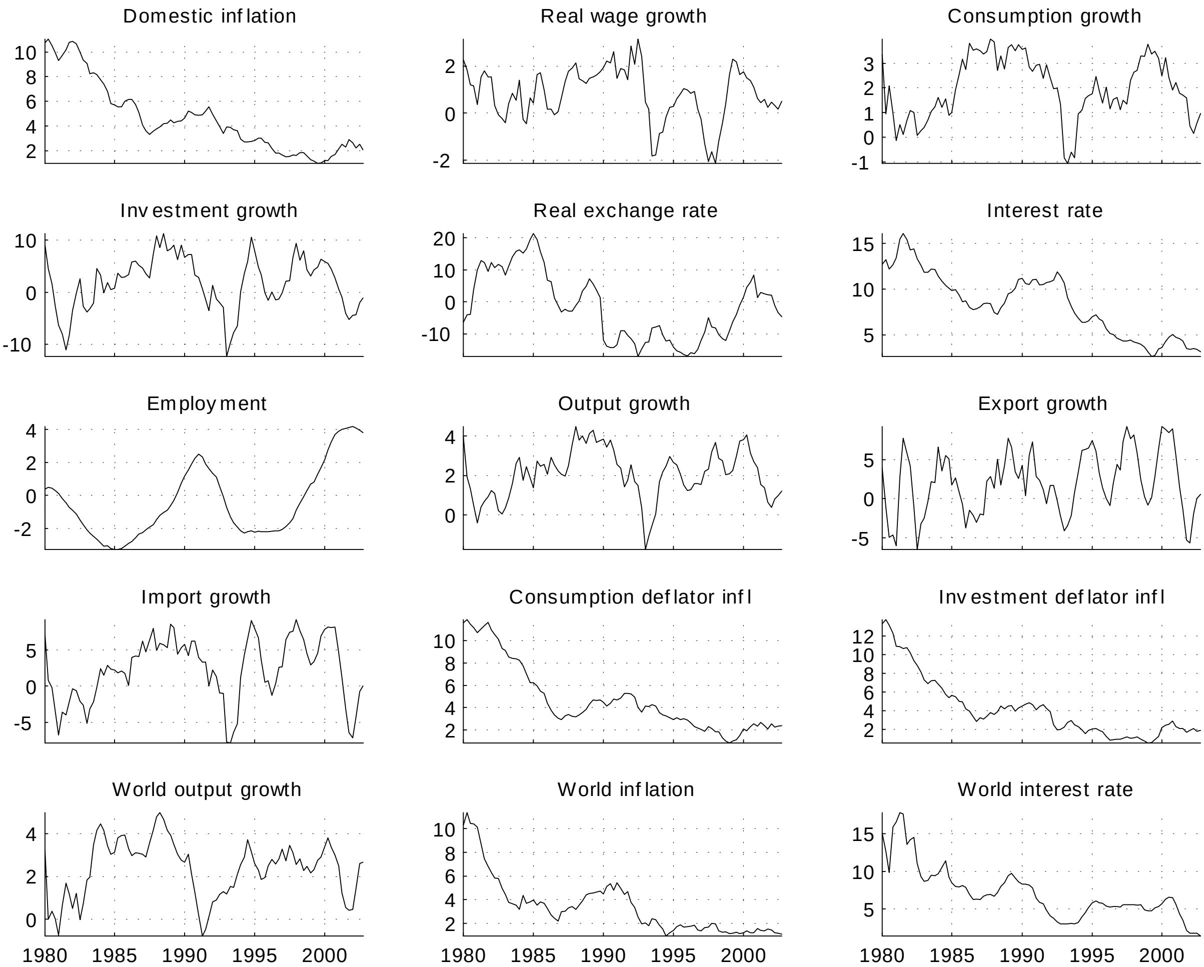

FIGURE 1. Euro area data 1980Q1-2002Q4, yearly growth rates.

To calculate the likelihood function of the observed variables we apply the Kalman filter. Second, draws from the joint posterior are generated using the Metropolis-Hastings algorithm (see Schorfheide (2000) for details).[5] In Table 2 we report the posterior mode estimates of the parameters.

[5]A posterior sample of 500, 000 post burn-in draws was generated. Convergence was checked using standard diagnostics such as CUSUM and ANOVA on parallel simulation sequences.

In order to further look into the connection between the marginal likelihoods and out-of-sample performance we compare four different specifications of the DSGE model. Apart from the benchmark model we also report estimation results in Table 2 for the following model specifications: *i*) with variable capital utilization, *ii*) with persistent domestic markup shocks, and *iii*) with IID markup shocks. We have chosen these specifications since a high or a low cost of varying the capital utilization (captured in a high or low value of the parameter σ_a) has rather large effects on the impulse response functions. For example, with variable capital utilization, marginal cost is smoother after a monetary policy shock which in turn also makes the response of inflation more smooth. For case *ii*) we find that allowing for persistent domestic markup shocks implies that the domestic price stickiness is estimated to a much lower number, see Table 2. Similarly, if all markup shocks are assumed to be independently distributed, the source of variation as well as the price stickiness parameters (ξ:s) are completely different. We interpret this as that the model needs either a high degree of price stickiness or highly correlated markup shocks to explain the high inflation inertia seen in the data. We also find a larger role for indexation to past inflation in this case, so that when less of the persistence is generated by correlated shocks there must be a larger role for intrinsic persistence (i.e. lagged inflation) to account for the inflation dynamics. Note that the alternative specifications are estimated using the data in first differences.

Figures 2 and 3 show the sequential estimates (posterior mode) of the different DSGE models' parameters when extending the data set year-by-year from 1994 and onwards. For each specification of the model most of the parameters appear to be relatively stable over time which is encouraging given that the parameters are updated according to this scheme in the subsequent rolling forecast evaluation. However, the model estimated with cointegration constraints show somewhat less stability. First of all, there is negative correlation between the habit formation parameter and the persistence of the consumption preference shocks. Second, the parameters related to investment (investment adjustment costs and the persistence and standard deviation of the investment-specific technology shock) are correlated over time and unstable. The instability in some of the estimates in the model estimated with cointegration imposed on the data is probably an effect of the rather large persistent movements in the cointegrating relations, in combination with a relatively short sample period (the shortest is 1980Q1-1993Q4 and the longest 1980Q1-2001Q4). All in all, it seems reasonable to start the evaluation of the forecasts as early as 1994, which leaves us with a relatively large evaluation sample.

3. Alternative forecasting models

The DSGE model is compared to several vector autoregressive (VAR) models, using both maximum likelihood estimates of the parameters and Bayesian posterior distributions. In addition, naïve forecasts based on both univariate random walks and the means of the most recent data observations are calculated.

The VAR systems consist of either seven or thirteen variables, with trending variables modelled in first differences. The first is a closed economy specification composed of the seven domestic variables: the domestic inflation rate, the short-run interest rate, employment, consumption, investment, GDP, and the real wage. The second is an open economy specification which additionally includes exports, imports, the real exchange rate, foreign inflation, the foreign interest rate, and foreign output. Note that the consumption and investment deflators have been excluded in the VARs for reasons of parsimony. Estimating the VARs in first

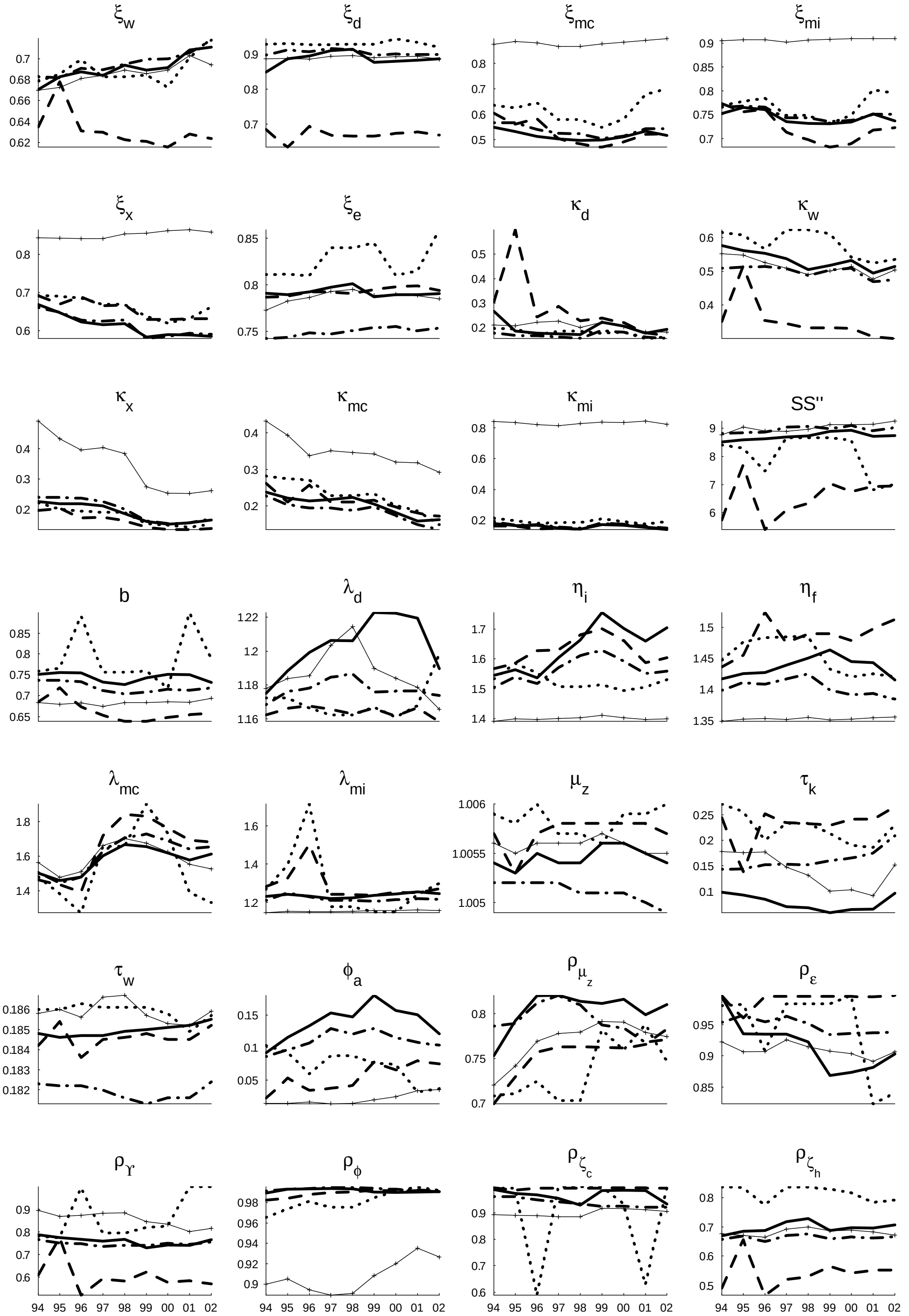

FIGURE 2. Sequential posterior mode estimates of the DSGE models' parameters using a year-by-year extended data set. DSGE diff. (—), DSGE coint. (···), Corr. mkup. (- - -), Var. cap. util. (– · –) and IID Markup (–+–).

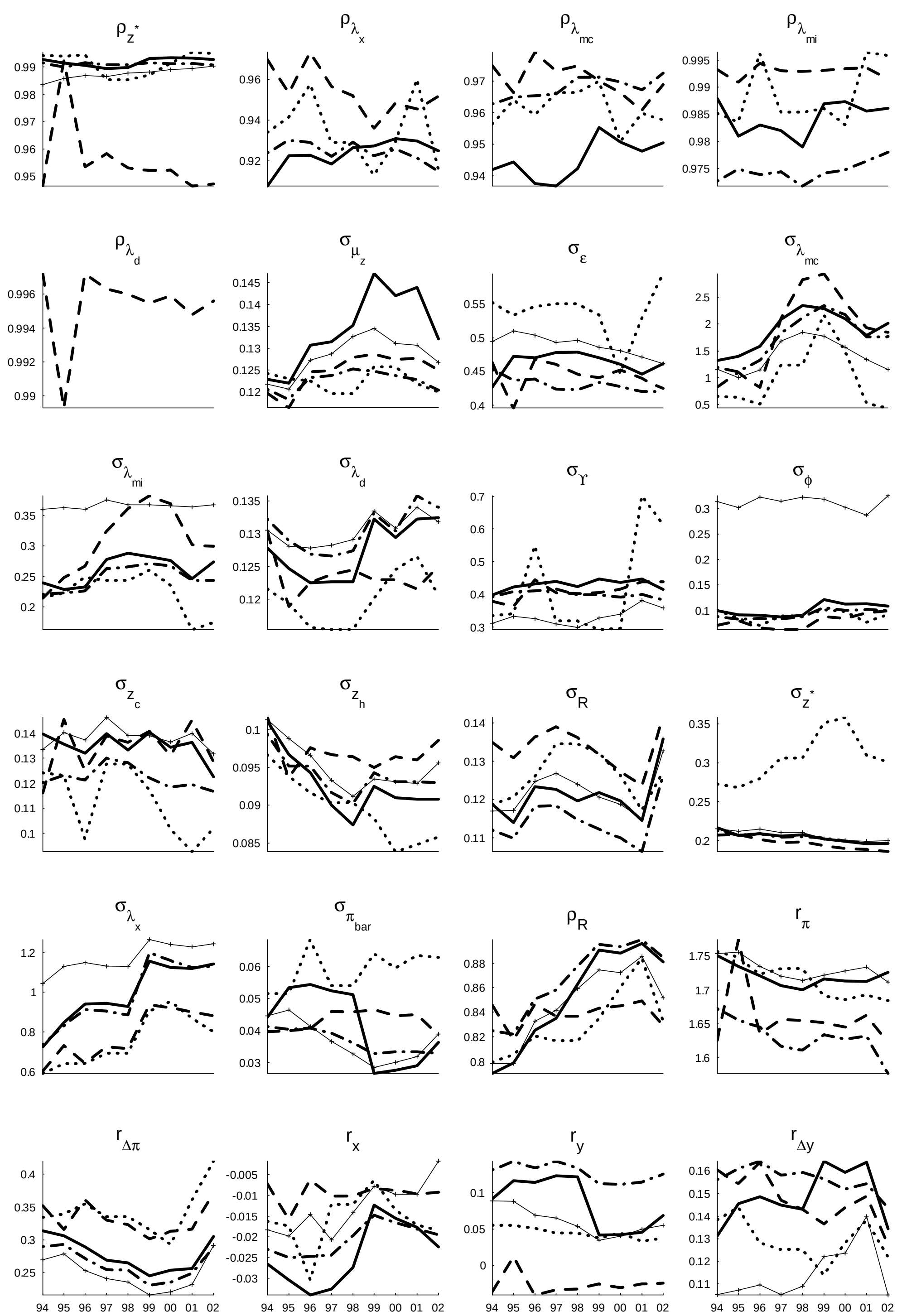

FIGURE 3. Sequential posterior mode estimates of the DSGE models' parameters using a year-by-year extended data set. DSGE diff. (—), DSGE coint. (···), Corr. mkup. (- - -), Var. cap. util. (– · –) and IID Markup (–+–).

differences may suffer from misspecification since we do not allow for any cointegration vectors. However, if the cointegrating relations are not stable over time, differencing may play a robustifying role, see e.g. Clements and Hendry (1998).

The usual parametrization of the VAR model reads

$$\Pi(L)x_t = \Phi d_t + \varepsilon_t, \tag{3.1}$$

where x_t is a p-dimensional vector of time series, $\Pi(L) = I_p - \Pi_1 L - ... - \Pi_k L^k$, and L the usual back-shift operator with the property $Lx_t = x_{t-1}$. The disturbances $\varepsilon_t \sim N_p(0, \Sigma)$, $t = 1, ..., T$, are assumed to be independent across time. $d_t = (1, d_{MP,t})'$ is a vector of deterministic variables. As noted in Section 2, the DSGE model embodies a time-varying inflation target which enables it to capture the downward shift in the nominal variables over the sample period. A regime dummy

$$d_{MP,t} = \begin{cases} 1 \text{ if } t \leq t^* \\ 0 \text{ if } t > t^* \end{cases}.$$

is included in the VARs as a proxy for this change in monetary policy. The date of the regime shift, t^*, is set to $1992Q4$ based on the posterior distribution of t^* presented in Villani (2005).

We will also consider an alternative parametrization of the VAR model of the form

$$\Pi(L)(x_t - \Psi d_t) = \varepsilon_t. \tag{3.2}$$

This somewhat non-standard parametrization of the VAR model in (3.2) is non-linear in its parameters, but has the advantage that the unconditional mean, or steady state, of the process is directly specified by Ψ as $\mathrm{E}_0(x_t) = \Psi d_t$. This allows us to put the BVAR and DSGE models more on par by using a prior on the steady state of the BVAR which is comparable to the steady state prior used in the DSGE models. To formulate a prior on Ψ, note that the specification of d_t implies the following parametrization of the steady state

$$\mathrm{E}_0(x_t) = \begin{cases} \psi_1 + \psi_2 & \text{if } t \leq 1992Q4 \\ \psi_1 & \text{if } t > 1992Q4 \end{cases},$$

where ψ_i is the ith column of Ψ. The prior on ψ_1 thus determines the steady state in the latter regime. The elements in Ψ are assumed to be independent and normally distributed *a priori*. The 95% prior probability intervals for the yearly steady state growth rates are given in Table 3. We will refer to specifications (3.1) and (3.2) as the BVAR and MBVAR (mean-adjusted Bayesian VAR), respectively.

The prior proposed by Litterman (1986) will be used on the dynamic coefficients in Π, with the following default values on the hyperparameters: overall tightness is set to 0.3, cross-equation tightness to 0.2 and a harmonic lag decay with a hyperparameter equal to one. See Litterman (1986) and Doan (1992) for details. Litterman's prior was designed for data in levels and has the effect of shrinking the process toward the univariate random walk model. We therefore modify the original Litterman prior by setting the prior mean on the first own lag to zero for all variables in growth rates. The two interest rates, employment and the real exchange rate are assigned a prior which centers on the AR(1) process with a dynamic coefficient equal to 0.9. In all VAR models we impose the small open economy restriction that the foreign variables are exogenously given, i.e., block exogenity of (π_t^*, y_t^*, R_t^*). Finally, the usual non-informative prior $|\Sigma|^{-(p+1)/2}$ is used for Σ.

Table 3: 95% prior probability intervals of Ψ

	π	$\triangle w$	$\triangle c$	$\triangle i$	R
ψ_1	$(1.54, 2.33)$	$(2.02, 2.83)$	$(2.02, 2.83)$	$(2.02, 2.83)$	$(4.93, 6.39)$
ψ_2	$(4, 7)$	$(-0.05, 0.05)$	$(-0.05, 0.05)$	$(-0.05, 0.05)$	$(3, 5)$
	$\widehat{E}$	Δy	x	$\Delta\widetilde{X}$	$\Delta\widetilde{M}$
ψ_1	$(-10, 10)$	$(2.02, 2.83)$	$(-10, 10)$	$(2.02, 2.83)$	$(2.02, 2.83)$
ψ_2	$(-0.05, 0.05)$	$(-0.05, 0.05)$	$(-0.05, 0.05)$	$(-0.05, 0.05)$	$(-0.05, 0.05)$
	Δy^*	π^*	R^*		
ψ_1	$(2.02, 2.83)$	$(1.54, 2.33)$	$(4.93, 6.39)$		
ψ_2	$(-0.05, 0.05)$	$(4, 7)$	$(3, 5)$		

Note: The prior on the steady state is specified in terms of yearly rates for the domestic and foreign inflation and interest rates (π, R, π^*, R^*) and in yearly growth rates for all real variables except employment and the real exchange rate (i.e., Δw, Δc, Δi, Δy, $\Delta\widetilde{X}$, $\Delta\widetilde{M}$, and Δy^*). For employment and the real exchange rate the prior is specified as deviations around the steady state.

The posterior distribution of the model parameters and the forecast distribution of the endogenous variables were computed numerically using the Gibbs sampling algorithm in Kadiyala and Karlsson (1997) for the parameterization in (3.1) and the Gibbs sampler in Villani (2005) for the specification in (3.2).

To sum up, we analyze two different VAR-systems (7 and 13 variables) with 1 to 4 lags. For each of these models, we employ two different specifications of the deterministic part of the process, given by eq. (3.1) and eq. (3.2), respectively. In addition to this we also estimate the 7- and 13-variables system with maximum likelihood. To save space we choose only to report the results from the VARs and BVARs with four lags. However, the forecasting results are similar across lag-lengths, possibly with a slight advantage for just the four lag models.

4. Measuring forecast accuracy

4.1. **The rolling forecast evaluation scheme.** We will analyze the out-of-sample precision of forecasts from the competing models in detail. A large set of accuracy measures will be employed to summarize the performance of the point forecasts as well as other aspects of the forecast distribution, such as the forecast uncertainty intervals. The performance of the forecasting models will be assessed using a standard rolling forecast procedure where the models' parameters are estimated using data up to a specified time period T where the dynamic forecast distribution of $x_{T+1}, ..., x_{T+h}$ is computed. The estimation sample is then extended to include the observed data at time $T+1$ and the dynamic forecast distribution of $x_{T+2}, ..., x_{T+h+1}$ is computed. This is prolonged until no data are longer available to evaluate the one-step ahead forecast. Notice that the BVARs are re-estimated at a quarterly frequency while the DSGE models are re-estimated only yearly. We start the rolling forecasts in $1993Q4$, with the first out-of-sample forecast produced for $1994Q1$. The final estimation period is $2002Q3$ which provides one 1-step ahead forecast to be evaluated against the final data point in our sample which is dated $2002Q4$. We consider the forecast horizons 1 to 8 quarters ahead. This gives us 36 hold-out observations for the 1-step ahead forecast and 28 observations on the longest horizon.

4.2. **Marginal likelihood as a measure of predictive performance.** There is a connection between out-of-sample predictive performance and Bayesian model posterior probabilities. The posterior probability of a model is proportional to the prior probability of that model multiplied by its marginal likelihood, or prior predictive density:

$$p_0(x_1, ..., x_T) = \int p(x_1, ..., x_T|\theta)p_0(\theta)d\theta.$$

It is important to note that the marginal likelihood is a predictive density based on the prior distribution $p_0(\theta)$ as a summary of the parameter uncertainty; no data is consumed to estimate the parameters of the model when computing the marginal likelihood. This makes it possible to interpret the marginal likelihood as a measure of out-of-sample predictive performance, rather than in-sample fit. This has been noted by several researchers, beginning with Jeffreys (1961), but is most clearly formulated in Geweke (1999) using the decomposition

$$p_0(x_1, ..., x_T) = \prod_{\tau=0}^{q} p_{s_\tau}(x_{s_\tau+1}, ..., x_{s_{\tau+1}}),$$

where $0 = s_0 < s_1 < ... < s_{q+1} = T$ partitions the sample into disjoint subperiods. $p_{s_\tau}(x_{s_\tau+1}, ..., x_{s_{\tau+1}})$ is the predictive density of $x_{s_\tau+1}, ..., x_{s_{\tau+1}}$ conditional on data up to time s_τ:

$$p_{s_\tau}(x_{s_\tau+1}, ..., x_{s_{\tau+1}}) = \int p_{s_\tau}(x_{s_\tau+1}, ..., x_{s_{\tau+1}}|\theta)p_{s_\tau}(\theta)d\theta,$$

where $p_{s_\tau}(\theta)$ is the posterior distribution of θ based on data up to time s_τ. An illustrative special case is obtained by letting $s_\tau = \tau$ for $\tau = 0, 1, ..., T-1$. We then obtain a decomposition of the marginal likelihood in terms of one step-ahead predictive densities

$$p_0(x_1, ..., x_T) = p_0(x_1)p_1(x_2) \cdots p_{T-1}(x_T).$$

The close connection between the marginal likelihood and out-of-sample forecasting accuracy may seem to make traditional out-of-sample prediction exercises obsolete. We give three arguments for why this is not the case. First, the above stated decomposition reveals that the marginal likelihood gives weight to the forecasting accuracy early in the sample where the prior is dominating the posterior distribution. For example, the predictive score of the first observation is based on parameters drawn directly from the prior distribution. This is of course entirely in line with the logic of marginal likelihoods - it values the combination of the model *and* the prior. However, the user of a forecasting model is likely to be more interested in the expected future forecasting performance based on the posterior distribution available at the time of the forecast. Second, the marginal likelihood evaluates whole forecast paths from $T+1$ to $T+h$. It cannot detect that some models may produce mediocre forecasts at shorter horizons and at the same time be relatively accurate, in comparison to other models, at longer horizons. Put differently, the marginal likelihood cannot be decomposed into h-step ahead predictive densities, $p_T(x_{T+h})$. Third, the marginal likelihood measures forecasting accuracy by the predictive score, a precision measure which focuses on the system as a whole. The user may be more concerned with forecasting a subset of these variables, or in other performance measures.

We use four different specifications of the DSGE model to contrast the out-of-sample forecasting performance in a traditional rolling event forecasting exercise to the marginal likelihoods. It is of course also possible to include BVARs in the set of models for which model probabilities are computed, see e.g. Smets and Wouters (2004). It is well known that the posterior distribution over a collection of models can be sensitive to the choice of prior distribution, see e.g. Sims' (2003) discussion of Smets and Wouters (2003). This may not be a severe problem if the models under consideration have similar structure so that the models' priors are constructed in essentially the same way. This is not the case in the DSGE vs BVAR comparison. In the former model, the prior is elicitated on the structural parameters, using economic theory and available microdata, whereas the priors in BVARs are mostly based on purely statistical considerations as in e.g. the Litterman (1986) prior. Since the microfounded DSGE prior is very likely to be substantially different from the statistical BVAR prior, the marginal likelihoods of the two models may very well be radically different even if the two models are very similar. An interesting alternative is developed in Del Negro and Schorfheide (2004), which may be described in this context as a way to form a continuous path between the DSGE and BVAR priors. An application of this methodology is given in Del Negro et al. (2004).

4.3. **Measuring the accuracy of point forecasts.** Let $\hat{x}_{t+h|t}$ denote the h-step-ahead posterior median forecast of x_{t+h}, standing at time t, and define $e_t(h) = x_{t+h} - \hat{x}_{t+h|t}$ as the corresponding forecast error. We will consider the usual univariate measures of accuracy of point forecasts, the mean absolute forecast error (MAE) and the root mean squared forecast error (RMSE):

$$MAE_i(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} |e_{i,t}(h)|, \tag{4.1}$$

$$RMSE_i(h) = \sqrt{N_h^{-1} \sum_{t=T}^{T+N_h-1} e_{i,t}^2(h)}, \tag{4.2}$$

where $e_{i,t}(h)$ is the ith element of $e_t(h)$ and N_h denotes the number of evaluated h-step-ahead forecasts.

We also consider two multivariate measures of point forecast accuracy based on the scaled h-step-ahead Mean Squared Error (MSE) matrix

$$\Omega_M(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \tilde{e}_t(h)\tilde{e}_t'(h). \tag{4.3}$$

where $\tilde{e}_t(h) = M^{-1/2}e_t(h)$. The matrix M acts as a scaling matrix that accounts for the differing scales of the forecasted variables and for the fact that the time series may be more or less intrinsically predictable in absolute terms. Commonly used scalar valued multivariate measures of point forecast accuracy are the log determinant statistic $\ln|\Omega_M(h)|$ and the trace statistic $\text{tr}[\Omega_M(h)]$. Note the relations $\ln|\Omega_M(h)| = \ln|\Omega_I(h)| - \ln|M|$ and $\text{tr}[\Omega_M(h)] = \text{tr}[M^{-1}\Omega_I(h)]$, so that the log determinant statistic is invariant to the choice of scaling matrix, whereas the trace statistic is not. Because of this, and to simplify the interpretation of the trace statistic, we set M equal to a diagonal matrix with the sample variances of the time series based on data from $1993Q1 - 2002Q4$ as diagonal elements. With M equal to a diagonal matrix, the trace statistic reduces to a simple weighted average of the RMSEs of the individual series.

The convenient information reduction provided by scalar valued size measures of Ω (M and h is dropped from $\Omega_M(h)$ here for notational convenience) may of course also hide important information. A more detailed view is given by the singular value decomposition of Ω: $\Omega = V\Lambda V'$,where $V = (v_1, ..., v_k)$, $V'V = I_k$, is the matrix of eigenvectors, Λ=diag$(\lambda_1, \lambda_2, ..., \lambda_k)$ is the diagonal matrix with ordered eigenvalues $\lambda_1 \geq \lambda_2, ..., \geq \lambda_k$. The eigenvalues are the variances of the principal components $y_{i,t} = v_i'\tilde{e}_t$. λ_1 is thus the variance of the least predictable (maximal forecast error variance) linear combination of the time series and λ_k is the variance of the linear combination of the time series with smallest forecast error variance. Since $\ln|\Omega| = \sum_{i=1}^{k} \ln \lambda_i$ and $\mathrm{tr}(\Omega) = \sum_{i=1}^{k} \lambda_i$, it is clear that $\mathrm{tr}(\Omega)$ will to a large extent be determined by the forecasting performance of the least predictable dimensions (largest eigenvalues), whereas $\ln|\Omega| = \sum_{i=1}^{k} \ln \lambda_i$ also takes into account the most predictable dimensions (smallest eigenvalues), sometimes to the extent of being dominated by them. To see the latter point, note that as $\lambda_k \to 0$ we have $\mathrm{tr}(\Omega) \to \sum_{i=1}^{k-1} \lambda_i$, but $\ln|\Omega| \to -\infty$, for any values of λ_i, $i = 1, ..., k-1$. A variable with a large (squared) coefficient in the first principal component is thus a major contributor to $\mathrm{tr}(\Omega)$, and a variable with a prominent role in the last principal component contributes substantially to $\ln|\Omega|$, at least when λ_k is close to zero. Thus, to get a detailed view of the multivariate accuracy measures one may look at the square of the elements of the eigenvectors, v_i, in combination with the eigenvalues, see Section 5.1 for an application.

4.4. **Measuring the accuracy of density forecasts.** The measures of forecasting performance presented so far value the accuracy of the point forecasts. The Bayesian methodology employed here allows us to easily obtain the exact finite sample joint forecast distribution of the system. To evaluate the out-of-sample performance of the multivariate forecast density as a whole we employ the log predictive density score (LPDS). The log predictive density score (LPDS) of the h-step-ahead predictive density, standing at time t, is defined as

$$S_t(x_{t+h}) = -2 \log p_t(x_{t+h}),$$

where $p_t(x_{t+h})$ denotes the h-step-ahead forecast distribution of the k-dimensional data vector x_{t+h}, standing at time t. Under the assumption that $p_t(x_{t+h})$ is a normal distribution, the LPDS can be written

$$S_t(x_{t+h}) = k \log(2\pi) + \log \left|\Sigma_{t+h|t}\right| + (x_{t+h} - \bar{x}_{t+h|t})' \Sigma_{t+h|t}^{-1} (x_{t+h} - \bar{x}_{t+h|t}),$$

where $\bar{x}_{t+h|t}$ and $\Sigma_{t+h|t}$ are the posterior mean and covariance matrix of the h-step ahead forecast distribution, respectively, standing at time t.[6] We will report the average LPDS over the hold-out sample

$$S(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} S_t(x_{t+h}),$$

where N_h is the number of h-step-ahead rolling forecasts.

The predictive score measures the conformity of the observations to the predictive density as a whole. Another aspect of the predictive density is the forecast intervals. There are many ways to construct a forecast interval with predetermined coverage probability, e.g. highest posterior density (HPD) intervals. We shall here restrict attention to forecast intervals with equal tail probabilities. A forecast interval is said to be well calibrated if the long run relative frequency of realized observations included in the forecast interval equals the pre-specified coverage probability of the interval (Dawid, 1982).

Formally, define the sequence of hit indicators of an h-step-ahead forecast interval with coverage probability α as

$$I_t^{\alpha}(h) = \begin{cases} 1 \text{ if } x_t \in [L_t^{\alpha}(h), H_t^{\alpha}(h)] \\ 0 \text{ if } x_t \notin [L_t^{\alpha}(h), H_t^{\alpha}(h)] \end{cases}$$

[6]If the forecast distribution is not at least approximately normal it may be estimated by a kernel density estimator. This will be very computationally demanding if the dimension of x_t exceeds two or three.

where $L_t^\alpha(h), H_t^\alpha(h)$ are the lower and upper limits of the interval at time t. The relative frequency of interval hits in the evaluation sample, $\hat{\alpha}_h = N_h^{-1} \sum_{t=T}^{T+N_h-1} I_t^\alpha(h)$, may then be compared to the pre-specified coverage rate α.

A more detailed analysis can be made for the one step-ahead forecasts. In that case, the hit sequence from a correct forecast interval follows an *iid* Bernoulli process with success probability α. This characterization does not hold for $h > 1$ as the forecast errors are then no longer independent. Christoffersen (1998) suggests using asymptotic likelihood ratio tests to test the Bernoulli hypothesis against several alternatives. As a Bayesian alternative to these tests we compute posterior probabilities of the following three hypotheses

$$\begin{aligned} H_0 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \overset{iid}{\sim} Bern(\alpha) \\ H_1 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \overset{iid}{\sim} Bern(\pi) \\ H_2 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \sim Markov(\pi_{01}, \pi_{11}), \end{aligned} \tag{4.4}$$

where π in H_1 and π_{01}, π_{11} in H_2 are estimated freely. The notation $Markov(\pi_{01}, \pi_{11})$ is here used to denote a general two-state Markov chain with transition probabilities $\pi_{01} = \Pr(0 \to 1)$ and $\pi_{11} = \Pr(1 \to 1)$. If H_0 is supported, the forecast intervals are correct, both in terms of coverage and independence of interval hits. If data supports H_1, the hit indicators are independent, but do not generate the intended coverage α. A large posterior probability of H_2 suggests a violation of the independence property of the interval. Note that even if H_2 receives the largest posterior probability, the coverage of the interval may still be correct. Whether or not the interval has the correct coverage when the evidence is in favor of H_2 is indicated by the relative distribution of the remaining probability mass on H_0 and H_1.

The posterior probabilities of H_0, H_1 and H_2 are computed as follows. Let n_0 and n_1 denote the number of zeros and ones, respectively, in the hit sequence. Let further n_{ij} denote the number of transitions from state i to state j in the Markov chain under hypothesis H_2, so that for example n_{01} is the number of zeros in the sequence which are followed by ones. Assuming independent priors $\pi \sim Beta(\gamma, \delta)$ in H_1, $\pi_{01} \sim Beta(\gamma_{01}, \delta_{01})$ and $\pi_{11} \sim Beta(\gamma_{11}, \delta_{11})$ in H_2, the marginal likelihoods of the three hypotheses are easily shown to be

$$\begin{aligned} m_0 &= \alpha^{n_0}(1-\alpha)^{n_1} \\ m_1 &= \frac{B(n_0+\gamma, n_1+\delta)}{B(\gamma,\delta)} \\ m_2 &= \frac{B(n_{01}+\gamma_{01}, n_{00}+\delta_{01})B(n_{11}+\gamma_{11}, n_{10}+\delta_{11})}{B(\gamma_{01},\delta_{01})B(\gamma_{11},\delta_{11})}, \end{aligned}$$

where $B(\cdot,\cdot)$ is the Beta function. We will present results for uniform priors on π, π_{01} and π_{11}, i.e. we set $\gamma = \delta = \gamma_{01} = \delta_{01} = \gamma_{11} = \delta_{11} = 1$.

5. Empirical results

5.1. **Point forecasts.** Figure 4 shows the root mean squared forecast errors in yearly percentage terms at the 1 to 8 quarters horizon from the baseline DSGE model, two VAR systems (open and closed economy specifications), and two naïve setups (univariate random walks and the means of the eight most recent data observations). The mean absolute forecast errors give similar results and to save space we have chosen only to report the RMSEs. We see from the figure that the DSGE model does very well in terms of forecasts on the real exchange rate, exports and imports, at both short and long horizons, suggesting that the open-economy aspects of the DSGE model are satisfactory modeled. The DSGE model also seems to project consumption, employment, and the consumption deflator inflation very well. For output and

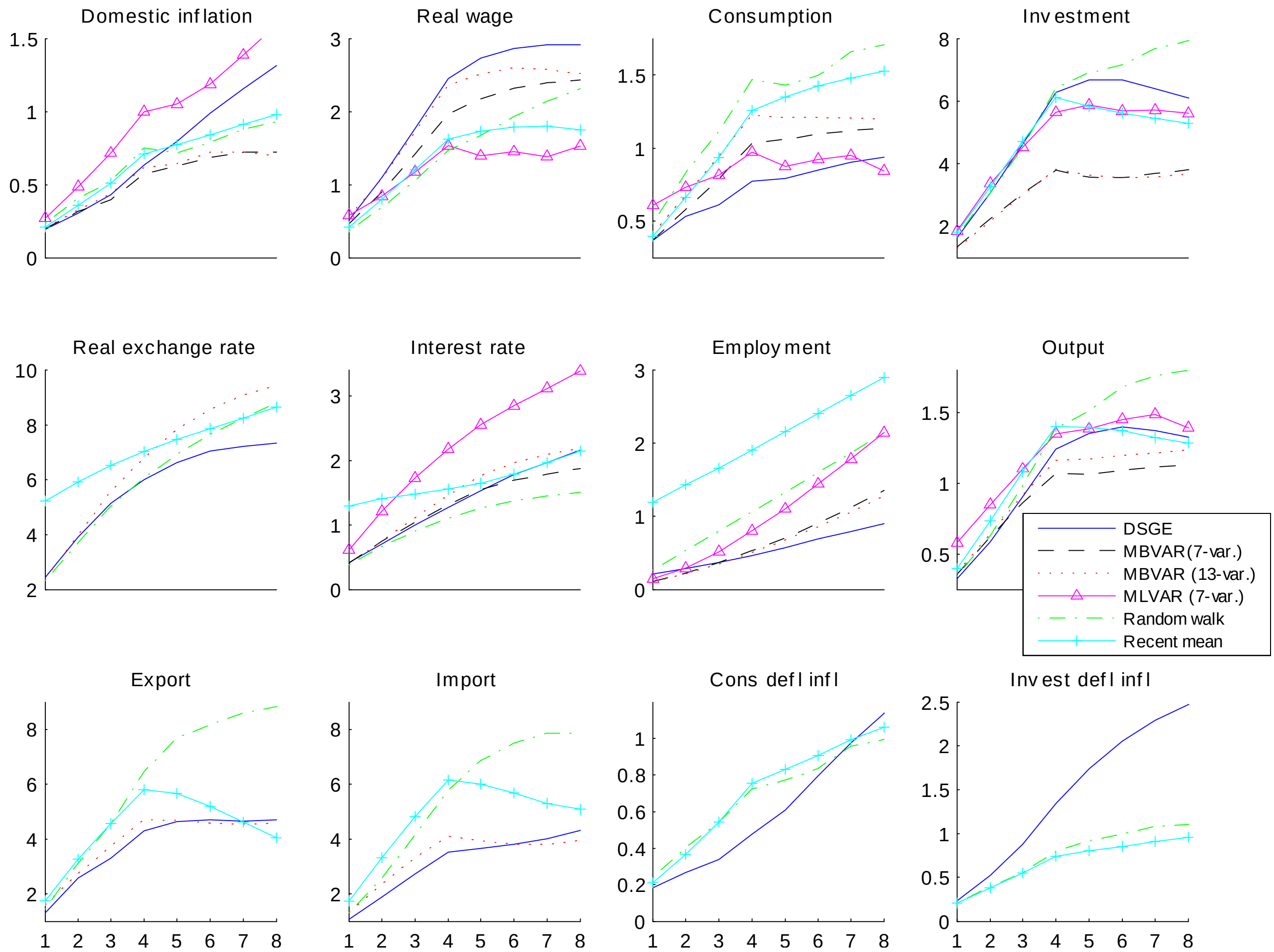

FIGURE 4. Root mean squared forecast errors for DSGE, BVARs, MLVARs and naïve.

domestic inflation the DSGE model does slightly better forecasts than the MBVARs at shorter horizons (1 and 2 quarters) but looses somewhat in the medium run. Note also that the one- and two-step-ahead forecasts from the DSGE model beat the random walk for most variables with the exception of the real wage and the investment deflator inflation. In addition, the DSGE model's forecasts outperform those of the MLVAR model on most variables and horizons (see Figures 4 and 6). However, at the eight quarter horizon the baseline DSGE model's forecast error for domestic inflation is a lot larger compared to the ones for the two Bayesian VAR systems. The DSGE model misses with about 1.3% on average while the forecast errors for domestic inflation in the MBVARs stay around 0.7%. It should be noted that the long-run properties of the DSGE and MBVARs are similar since the latter has a prior on the unconditional mean that is comparable to the steady state prior in the DSGE model. It is thus not obvious that the DSGE model's theoretical structure should matter more in the long run, and therefore have an advantage over the MBVARs in the forecasting performance at those particular horizons.

Figure 5 depicts the RMSEs for the four different specifications of the DSGE model estimated with data in first differences, together with the benchmark specification estimated with

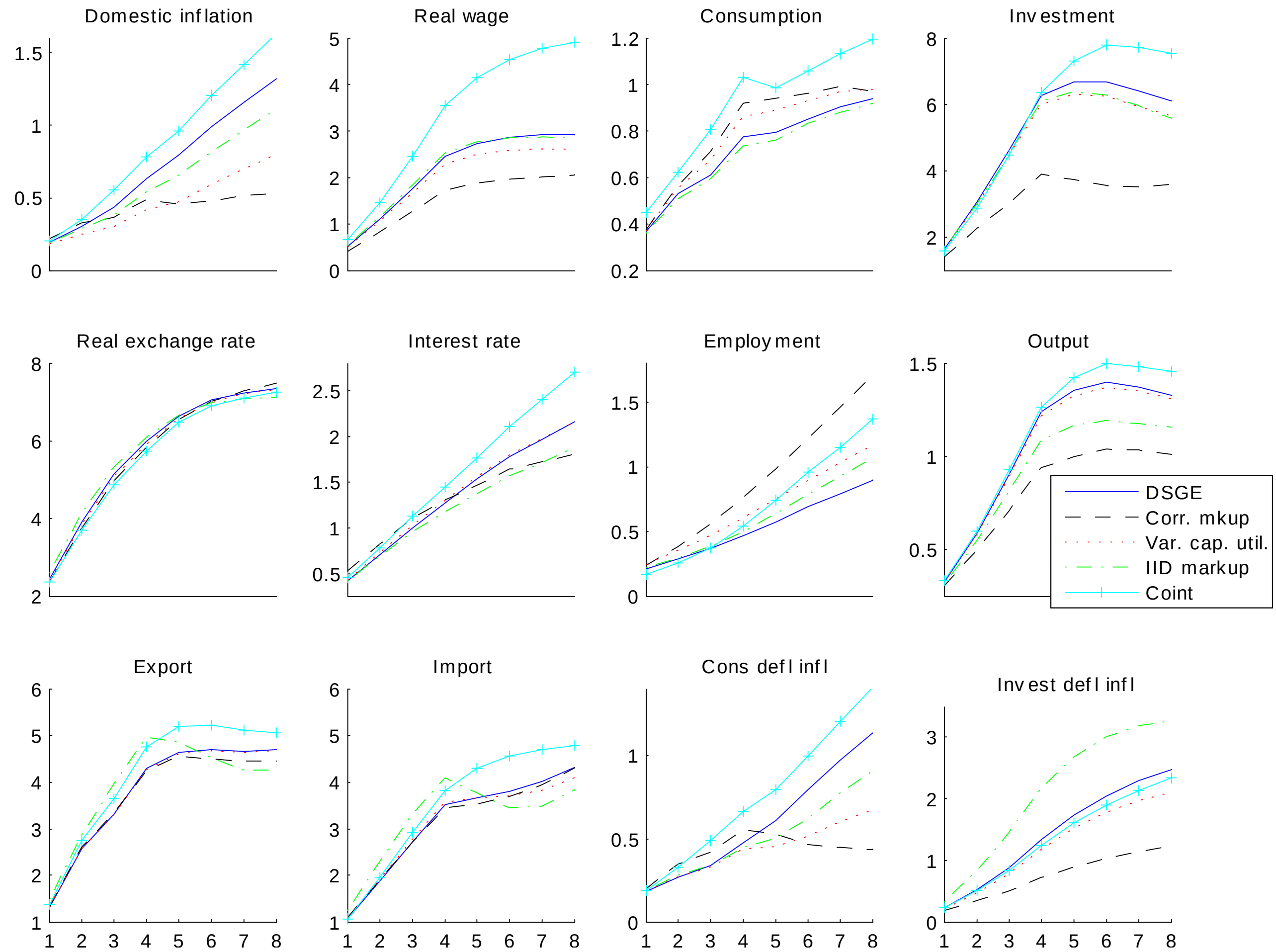

FIGURE 5. Root mean squared forecast errors for the DSGE models.

the DSGE models' cointegration restrictions imposed. The figure shows that the accuracy of the domestic inflation forecasts from the baseline DSGE model is a lot worse than the ones generated by the DSGE model with correlated markup shocks, which in turn is more in line with the MBVAR evidence (the BVARs and MLVAR models behave more similarly to the benchmark DSGE model). The problem is that the baseline model on average overpredicts both inflation and the real wage more often at longer horizons than the model with correlated markup shocks (not shown). By way of some simple experiments, we found that the main reason for this is the higher price stickiness parameter in the baseline DSGE which induces more inflation inertia than the model with correlated markup shocks. The baseline DSGE model consequently has more difficulties capturing upturns and downturns in the inflation series than, for example, the model with correlated markup shocks.[7]

Note also that imposing the models' cointegration restrictions in the estimation on the baseline specification leads to inferior forecasting performance on almost all variables and horizons (see Figure 5). One explanation for this behavior is that the cointegrating relations implied

[7]However, also other parameters contribute to the inflation persistence, such as a higher wage indexation and larger responses to the output gap in the monetary policy rule (cf. Table 2).

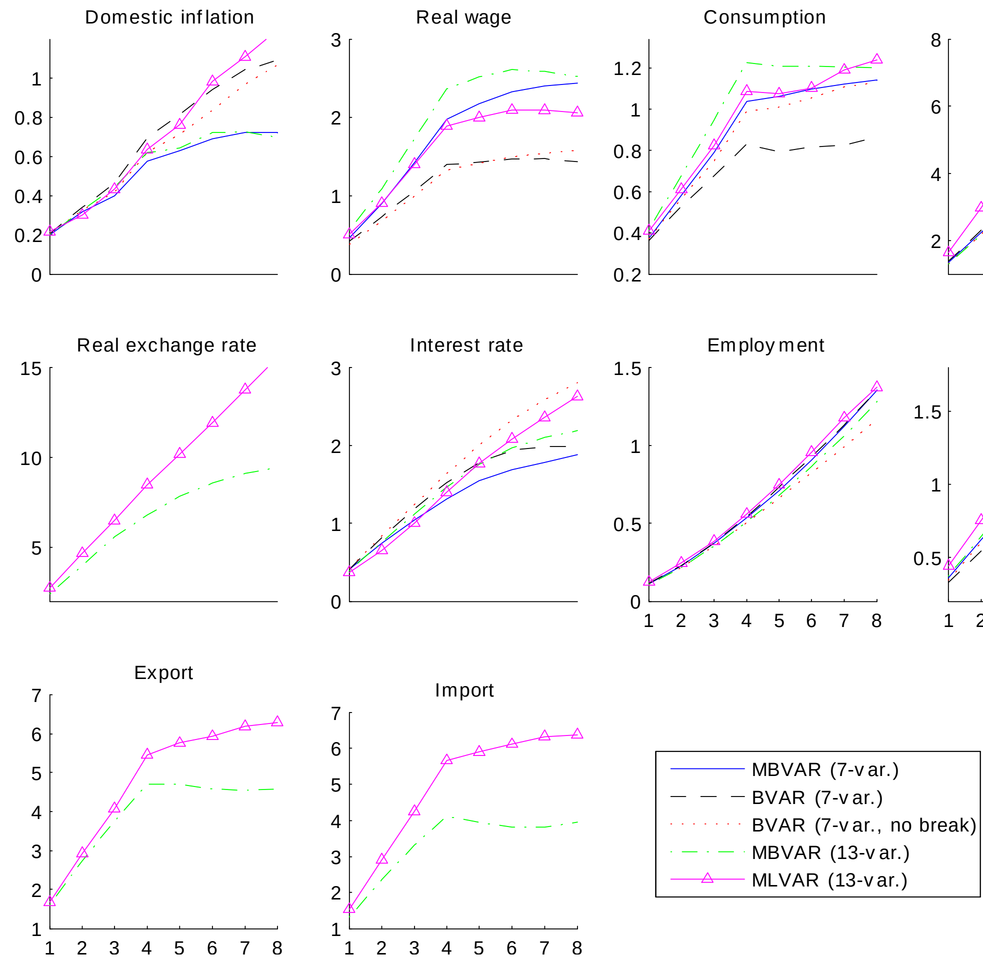

FIGURE 6. Root mean squared forecast errors for the VAR models.

by the DSGE model display a large degree of persistence during the sample period. In order to capture this feature of the data, the cointegration model is estimated to have both more intrinsic persistence (i.e., larger nominal frictions) and a higher correlation in the exogenous shock processes, compared to the baseline model estimated on data in first differences (see Table 2). This in turn causes the cointegration model to generate more persistent forecasts, which are generally not a feature of the actual outcome in the forecasting evaluation period where, for example, inflation is consistently low and less persistent than in the earlier part of the sample.

In Figure 6 we display the RMSEs for the various VAR systems. We see that the MBVARs with a prior on the steady state seems to do better in terms of the forecasts on inflation at longer horizons. On the other hand, the MBVAR models seem instead to perform worse on some of the real variables such as e.g. the real wage. The difference between the Bayesian VARs at longer horizons is to a large extent explained by the MBVARs' prior on the steady state in Table 3. The average inflation rate during the evaluation period turned out to be near the center of the steady state prior which explains the good long run forecasts of inflation from the MBVAR models. Likewise, the poor real wage forecasts are explained by the lack of correspondence between the realized real wage growth in the evaluation period and the steady state prior (compare Figure 1 and Table 3).

Table 4 shows the multivariate accuracy measures for the point forecasts, the log determinant and the trace of the MSE matrix (see equation (4.3)). In order to be able to compare the multivariate measures across the different models, we have in the first set chosen to include only the variables that are common to all models. The first set of multivariate measures are therefore based on the matrix of forecast errors from domestic inflation, the real wage, consumption, investment, employment, the interest rate and output. According to both the log determinant and the trace statistics, the BVAR models appear overall to have better accuracy on the one and four quarter ahead forecasts than the ones generated from the different DSGE specifications. However, at the 8 quarter horizon the forecasts from the DSGE model outperforms those of the BVARs, at least judging from the log determinant statistic.

The multivariate measures based on the domestic variables consequently seem to suggest that the DSGE model performs best at the 8 quarter horizon, while the univariate RMSEs on e.g. domestic inflation and the real wage indicate the opposite. The substantially worse performance of the DSGE models at the one quarter horizon is also hard to understand simply by looking at the univariate RMSEs. A closer look at the multivariate measures using the spectral decomposition of the MSE matrix discussed in Section 4.3 explains the seemingly incompatible results in Figures 4-6, on the one hand, and the multivariate measures in Table 4, on the other. Figure 7 displays the eigenvalues of the MSE matrix, both on original and log scale, at the $1, 4$ and 8 quarter horizon for four of the models. The log determinant statistic equals the sum of the log eigenvalues of the MSE matrix. It is therefore clear from the right column of Figure 7 that the large difference in forecasting performance between the DGSEs and BVARs captured by this statistic at the first quarter horizon is dominated by the smallest eigenvalue. The DSGE models inferior forecast performance at the one quarter horizon therefore comes from their inability to predict those variables which account for the major part of the last principal component at the shortest horizon. Looking at the subgraphs in the right column of Figure 8, which depicts the relative weight on the variables in the eigenvector with smallest eigenvalue (v_{jk}^2 for the jth variable), it is clear that this principal component at the shorter horizons is essentially the forecast errors of the employment series. The one-quarter ahead RMSEs of the employment series in Figure 4 are small for all models, but the *relative* difference between the DSGE models and the BVARs are substantial: the RMSE of employment at the first horizon in the benchmark DSGE is almost twice those of the two BVARs. Since the log determinant measure is very sensitive to the performance on the most predictable dimensions, this minor difference between the models receives a very large weight in the log determinant measure. Note also that the forecast errors of employment is still the driving force of the smallest eigenvalue at horizon 4 (Figure 8), but here the difference in log determinant statistic across models is no longer dominated by this eigenvalue (see Figure 7). At the four quarter horizon it is mainly the largest eigenvalue which is dominating the comparison. The determinants of this eigenvector are given in the left hand column of Figure 8. The relatively good multivariate performance of the DSGE model with correlated markup shocks and the seven variable BVAR is in part explained by the fairly large weight on real wage, a variable which these two models predict more accurately than the benchmark DSGE. Finally, on the eight quarter horizon the picture is more complicated, but the poor performance of the benchmark DSGE on the long run forecasts of the real wage (which drives the largest eigenvalue, see Figure 8) is more than compensated by its good forecasts of the variables contributing to the smallest eigenvalues.

Table 4: Multivariate accuracy measures

Hori-zon	Model	Domestic variables ($\pi_t, y_t, R_t, W_t, C_t, I_t, E_t$)			3-variables (π_t, y_t, R_t)			6-variables ($\pi_t, y_t, R_t, x_t, \tilde{X}_t, \tilde{M}_t$)		
		Log determinant statistic	Trace statistic	Log predictive density score	Log determinant statistic	Trace statistic	Log predictive density score	Log determinant statistic	Trace statistic	Log predictive density score
1Q	DSGE, Benchmark	-21.470	0.760	*7.114*	*-8.776*	*0.422*	**1.534**	<u>-17.657</u>	<u>0.657</u>	<u>12.589</u>
	DSGE, with variable capital utilization	-21.224	0.757	8.150	**-8.899**	<u>0.420</u>	<u>1.583</u>	**-17.823**	**0.654**	*12.649*
	DSGE, correlated markup shocks	-21.597	0.726	7.509	-8.142	0.466	2.332	-16.858	0.684	13.639
	DSGE, all markup shocks iid	-21.121	0.770	8.391	<u>-8.853</u>	**0.415**	1.680	-16.423	0.693	13.794
	DSGE, cointegration	-21.851	0.876	7.442	-8.544	0.442	*1.656*	*-17.412*	*0.669*	**12.545**
	7-variables BVAR (mean adjusted)	<u>-23.629</u>	*0.720*	8.235	-8.576	0.439	2.606			
	7-variables BVAR (standard)	*-23.526*	<u>0.692</u>	<u>5.517</u>	-8.670	0.430	2.172			
	7-variables BVAR (no break/standard)	**-23.787**	**0.679**	**5.146**	-8.637	0.439	2.172			
	13-variables BVAR (mean adjusted)	-23.115	0.785	9.001	-8.661	0.443	3.341	-17.330	0.716	14.605
	7-variables VAR (ML)	-19.429	1.035		-5.925	0.659				
	13-variables VAR (ML)	-22.389	0.803		-8.150	0.484		-16.141	0.787	
	Random walk	-20.158	0.765		-8.327	0.451		-16.468	0.697	
4Q	DSGE, Benchmark	-8.703	2.920	22.949	-2.494	1.433	*8.007*	<u>-7.134</u>	*2.051*	<u>24.841</u>
	DSGE, with variable capital utilization	-8.556	2.754	23.968	**-3.303**	*1.297*	**7.594**	**-7.446**	<u>1.955</u>	**24.564**
	DSGE, correlated markup shocks	*-9.127*	*2.223*	23.101	*-2.645*	**1.196**	9.252	-6.470	**1.873**	26.623
	DSGE, all markup shocks iid	-8.303	2.874	25.011	<u>-2.895</u>	<u>1.266</u>	<u>7.792</u>	-4.952	2.067	26.385
	DSGE, cointegration	<u>-9.200</u>	3.799	*22.523*	-2.357	1.600	8.342	*-7.057*	2.227	*25.275*
	7-variables BVAR (mean adjusted)	**-9.668**	2.454	<u>21.915</u>	-2.120	1.318	9.071			
	7-variables BVAR (standard)	-8.395	**2.140**	22.674	-1.672	1.405	9.095			
	7-variables BVAR (no break/standard)	-8.946	<u>2.180</u>	**21.434**	-1.850	1.467	9.181			
	13-variables BVAR (mean adjusted)	-8.309	2.820	24.707	-1.955	1.441	9.718	-5.387	2.191	26.235
	7-variables VAR (ML)	-5.943	2.798		0.809	2.030				
	13-variables VAR (ML)	-7.154	2.712		-1.414	1.571		-4.302	2.578	
	Random walk	-2.996	2.689		-1.221	1.558		-2.533	2.571	
8Q	DSGE, Benchmark	<u>-8.122</u>	3.720	**26.667**	-1.202	2.287	*10.114*	-6.024	2.855	<u>27.802</u>
	DSGE, with variable capital utilization	-7.838	3.279	27.345	<u>-1.665</u>	1.863	**9.547**	<u>-6.721</u>	<u>2.514</u>	**27.334**
	DSGE, correlated markup shocks	**-8.568**	<u>2.624</u>	27.703	**-2.046**	**1.434**	10.694	**-7.274**	**2.214**	29.021
	DSGE, all markup shocks iid	*-7.859*	3.464	27.385	*-1.284*	1.962	<u>9.636</u>	*-6.226*	*2.521*	*27.954*
	DSGE, cointegration	-6.831	5.437	28.077	-0.698	2.775	10.779	-5.277	3.309	28.604
	7-variables BVAR (mean adjusted)	-7.651	*2.982*	*26.840*	-0.916	<u>1.633</u>	10.722			
	7-variables BVAR (standard)	-5.622	**2.583**	26.899	0.188	1.932	11.140			
	7-variables BVAR (no break/standard)	-6.180	2.916	<u>26.697</u>	0.790	2.227	11.743			
	13-variables BVAR (mean adjusted)	-6.383	3.122	30.631	-0.308	*1.775*	11.345	-3.956	2.570	28.376
	7-variables VAR (ML)	-3.014	3.576		2.073	2.941				
	13-variables VAR (ML)	-3.564	3.610		0.444	2.541		-1.841	3.845	
	Random walk	0.134	3.618		-0.253	2.003		-0.773	3.462	

Note: Bold, underlined, and italicized numbers indicate the first, second and third best forecasting model for each measure.

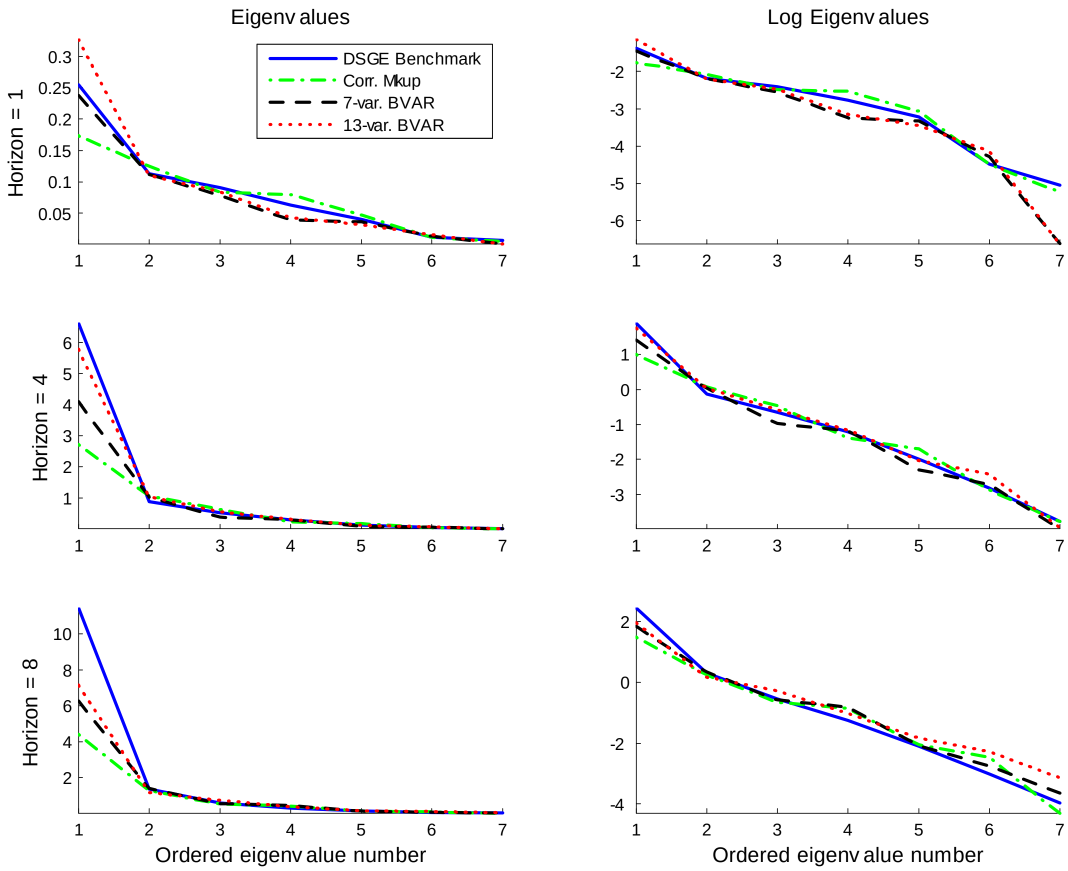

FIGURE 7. Eigenvalues (left column) and log of eigenvalues (right column) of the MSE matrix at different horizons.

The eigenvalues on the original scale in the left column of Figure 7 show that the trace statistic is mostly determined by the largest eigenvalue, at least at the longer horizons. The decomposition of this eigenvalue in the left column of Figure 8 may similarly be used to investigate the differences in the trace statistic.

Since the multivariate measures run the risk of being dominated by a specific variable which may be of minor interest (e.g., employment), Table 4 also shows the log determinant and trace of the MSE matrix from two other sets of variables. One is based on the forecast errors of domestic inflation, output and the interest rate, and another on these three variables together with the real exchange rate, exports and imports. Looking only at the three domestic variables it appears as if the DSGE models have a better chance of replicating the forecasting performance of the BVARs also at shorter horizons. The same holds true when adding the performance in terms of the open economy variables (i.e., the real exchange rate, exports and imports) to these three variables. Both the log determinant and the trace statistic indicate that the DSGE models have better forecast accuracy for this set of variables on the 1, 4 and 8 quarter horizons.

Turning to the comparison between the MBVAR and the BVAR without a prior on the unconditional mean, we find that the MBVAR does a good job in capturing the joint accuracy of the forecasted variables, at least according to the log determinant.

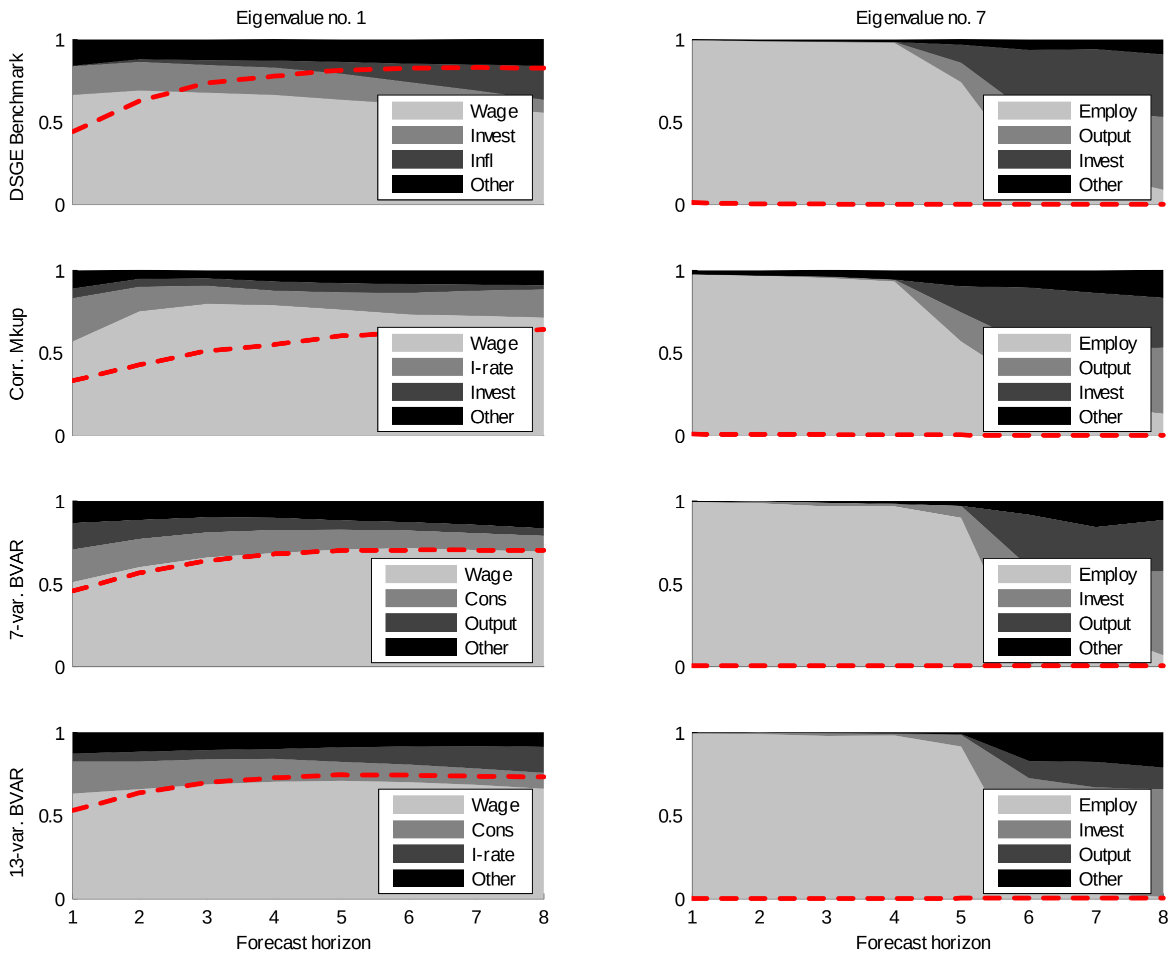

FIGURE 8. Relative contribution to the smallest and largest eigenvalue of the MSE matrix (square of the elements in the eigenvector). The superimposed red dashed line depicts the perecntage of total variation explained by the eigenvector, $\lambda_i / \sum_{j=1}^{k} \lambda_j$ where λ_i is the ith largest eigenvalue.

5.2. **Density forecasts.** Figure 9 and 10 show the accuracy of the one and four quarter ahead forecast intervals in terms of the empirical coverage probabilities for the baseline DSGE model and three BVAR specifications. The horizontal axis depicts the (intended) coverage probability of the interval and the vertical axis the empirical coverage rate obtained in the hold-out sample. This is measured from the sequence of hit indicators which determines how often a certain forecast interval (say for example 75% density) covers the actual data observations at the hth horizon. A good model should of course be equipped with a forecast density that is in close correspondence with the actual coverage probability, that is the empirical coverage rate should be located on the 45 degree line. The one-step ahead empirical coverage probabilities is based on 36 hit indicators, while the four-step ahead empirical coverage probability is calculated from 32 observations. The uncertainty in estimating percentiles from little more than 30 observations is of course large, especially in the tail of the distribution, and the exact numbers in Figure 9 and 10 should not be over-emphasized. The empirical coverage probabilities of the forecast intervals for the DSGE model seem in general to be more balanced at the one quarter horizon than the ones for the BVARs. This is especially true for domestic inflation and employment, where the BVAR forecast intervals' are too wide. At the four quarter horizon (Figure 10), the DSGE seems to deteriorate in comparison to the BVARs. A reason for the worse properties of the DSGE model could be the internal propagation of the disturbances hitting the economy. The processes for the disturbances are generally highly correlated, which implies that the uncertainty induced by these shocks amplify over the horizon and generate wider uncertainty bands.

Table 4 reports the log predictive density score. Once again we have in the first set chosen to include only the subset of variables that are common to the different models we evaluate, i.e., domestic inflation, the real wage, consumption, investment, employment, the interest rate and output. The LPDS based on these variables appear to suggest a better overall forecast density for the BVARs, at least at the shorter horizons, while the DSGE models gain ground on the longest horizon. However, the LPDS from the joint forecasts of only domestic inflation, output and the interest rate indicate that the DSGE models have better forecasting accuracy than the BVARs also at the one- and four-quarter horizons. Additionally, extending the second set of variables with the real exchange rate, imports and exports, we find that almost all DSGE specifications outperform the 13-variable MBVAR on all horizons. It is thus of crucial importance to acknowledge that also the LPDS can be dominated by variables that the forecaster is less interested in.

It is interesting to compare the LPDS of the seven domestic variables in the four DSGE model to the marginal likelihood of those four models in Table 2. The marginal likelihoods imply the following posterior probabilities: 0.9976, 0.0003, 0.002 and 0 for the benchmark model and the models with variable capital utilization, persistent markup shocks or iid markup shocks, respectively.[8] The model probabilties thus gives an overwhelming support to the baseline DSGE model with no variable capital utilization. It should be noted that the marginal likelihoods are evaluated on all 15 variables, whereas the LPDS presented in Table 4 only values the forecast density of the seven domestic variables. Comparing the marginal likelihoods of the DSGE models to the LPDS for the set of seven domestic variables in Table 4, we nevertheless see an exact correspondence in the ranking of DSGE models at the 1 and 4 horizons. The benchmark DSGE has the smallest LPDS also at the eight quarter horizon, but the ranking of the remaining three models is in conflict with the ranking based on the marginal likelihood.

The marginal likelihood is sometimes somewhat sloppily referred to as a measure of forecasting performance. This is problematic as forecasting performance is for most people mainly connected to point forecasts, while any connection between the marginal likelihood and out-of-sample forecast performance goes through the density forecasts, see Section 4.2. Judging by the evidence presented here, the use of the marginal likelihood as a convenient summary of the out-of-sample performance of point forecasts is highly questionable. For example, all four models generate more or less the same RMSE for variables such as domestic inflation, output, the real exchange rate, exports and imports at the short horizons (see Figure 5). The difference between the best and the worst model's one-step-ahead forecasts for both inflation and output is less than 0.03 percent. At longer horizons the differences between the models are larger, and perhaps surprisingly the baseline model generates the lowest RMSE only for one variable, employment, at the eight quarter horizon. Also the multivariate point forecast accuracy measures indicate a loose connection between marginal likelihood and out-of-sample

[8]It should be noted that these differences in posterior odds remain fairly constant over the evaluation period, at least to the degree of keeping the ranking of the models unaltered. Note also that the marginal likelihood of the DSGE model estimated on data where cointegration restrictions have been imposed cannot be directly compared to the other DSGE specifications.

performance. There is a slight (forecasting) edge for the model with correlated markup shocks, at all horizons, according to both the log determinant and the trace of the multivariate squared forecast error matrix for the domestic variables (see Table 3). The model with lower posterior probability thus outperforms the baseline model with higher posterior probability in terms of the point forecast accuracy.

The joint hypothesis test of correct interval coverage and independence of one-step ahead forecast interval hits (see Section 4.4) is presented in Table 5. The table shows the posterior probabilities of the three models in equation (4.4) for the 70% forecast interval. A well-calibrated forecasting model, where the interval hits are independent and the interval has the intended coverage, implies H_0 as the correct hypothesis . From the table follows that the baseline DSGE model has most probability mass on H_0. From Table 5 also follows that the benchmark DSGE model has somewhat better calibrated forecast intervals than the other DSGE specifications. As mentioned above, the model with correlated markup shocks does a lot better in terms of the point forecast accuracy of domestic inflation at longer horizons but on 1 and 2 quarters ahead the inflation forecast accuracy in the different DSGE specifications are about the same. However, from Figure 11 follows that the inflation forecast intervals are a lot wider in the model with correlated markups than in the baseline DSGE model. The empirical coverage rate is hence too large.

6. Conclusions

This paper has evaluated the forecasting performance of an open economy dynamic stochastic general equilibrium model for the Euro area against a wide range of reduced form forecasting models such as VARs, BVARs, univariate random walks and naïve forecasts based on the means of the most recent data observations.

The DSGE model performs very well in terms of univariate point forecasts on the open economy variables such as the real exchange rate, exports and imports. The RMSEs speak in favour of the DSGE model for these variables at both long and short horizons, suggesting that the open economy aspects are reasonably modeled. In terms of the domestic variables, the DSGE model also seems to forecast output, consumption and employment very well, but has some difficulty with the long run projections of domestic inflation and the real wage.

The multivariate point forecast accuracy measures, which take the joint forecasting performance of the domestic variables into account, indicate that the DSGE models give more accurate forecasts than the BVARs at the medium- to long-term horizons ($4-8$ quarters ahead). The advantage of the DSGE model is perhaps due to the richer theoretical structure that probably has a larger impact on the forecasts in the long-run, where the historical patterns captured in the VAR-systems can lead to more erroneous forecasts, at least without a prior on the steady state.

Turning to the overall density forecast accuracy, the differences between the models appear to be relatively small. Again, the baseline DSGE model seems to produce a somewhat better multivariate forecast density at longer horizons, while the BVARs have an overall forecasting advantage predominately at shorter horizons.

A caveat with the analysis in this paper is that we are using ex post data and not real-time data. The latter could perhaps change the ranking of the models, even if the same data are used in both the DSGE and the BVAR models. Another important issue for future work is to include more shocks with permanent effects in the model. The poor forecasting performance of the DSGE when imposing the model-implied cointegration properties suggests that it would be fruitful to incorporate more shocks with long-run effects than just the unit-root shock in total factor productivity considered here.

Table 5: Calibration inference for forecast intervals with a coverage probability of 70%

Horizon	Model		Inflation	Real wage	Consumption	Investment	Real exchange rate	Interest rate	Employment	Output	Export	Import	Consum. .deflator	Investment deflator
1Q	DSGE, Benchmark	H_0	0.713	0.697	0.677	0.634	0.760	0.222	0.084	0.435	0.011	0.149	0.498	0.011
		H_1	0.195	0.210	0.185	0.260	0.162	0.162	0.016	0.317	0.708	0.619	0.204	0.708
		H_2	0.092	0.092	0.138	0.106	0.079	0.616	0.900	0.248	0.282	0.232	0.298	0.282
	DSGE, with variable capital utilization	H_0	0.222	0.729	0.498	0.501	0.760	0.498	0.187	0.127	0.011	0.149	0.451	0.031
		H_1	0.162	0.167	0.204	0.365	0.162	0.204	0.056	0.527	0.708	0.619	0.329	0.441
		H_2	0.616	0.104	0.298	0.134	0.079	0.298	0.757	0.346	0.282	0.232	0.220	0.528
	DSGE, correlated markup shocks	H_0	0.201	0.498	0.222	0.631	0.618	0.514	0.160	0.049	0.011	0.138	0.149	0.001
		H_1	0.314	0.204	0.162	0.259	0.254	0.155	0.037	0.699	0.708	0.575	0.619	0.653
		H_2	0.485	0.298	0.616	0.110	0.128	0.331	0.804	0.252	0.282	0.287	0.232	0.346
	DSGE, all markup shocks iid	H_0	0.636	0.253	0.222	0.634	0.760	0.222	0.286	0.049	0.497	0.049	0.677	0.180
		H_1	0.261	0.352	0.162	0.260	0.162	0.162	0.065	0.699	0.362	0.699	0.185	0.525
		H_2	0.103	0.395	0.616	0.106	0.079	0.616	0.649	0.252	0.141	0.252	0.138	0.295
	7-variables BVAR	H_0	0.031	0.763	0.722	0.049		0.011	0.000	0.222				
	(mean adjusted)	H_1	0.441	0.146	0.198	0.699		0.708	0.500	0.162				
		H_2	0.528	0.091	0.080	0.252		0.282	0.500	0.616				
	7-variables BVAR	H_0	0.497	0.768	0.497	0.138		0.497	0.497	0.150				
	(standard)	H_1	0.362	0.164	0.362	0.575		0.362	0.362	0.621				
		H_2	0.141	0.068	0.141	0.287		0.141	0.141	0.230				
	7-variables BVAR	H_0	0.497	0.498	0.497	0.049		0.305	0.636	0.049				
	(no break/standard)	H_1	0.362	0.204	0.362	0.699		0.477	0.261	0.699				
		H_2	0.141	0.298	0.141	0.252		0.218	0.103	0.252				
	13-variables BVAR	H_0	0.004	0.046	0.194	0.050	0.049	0.001	0.000	0.222	0.722	0.201		
	(mean adjusted)	H_1	0.238	0.310	0.080	0.715	0.699	0.653	0.500	0.162	0.198	0.314		
		H_2	0.758	0.644	0.726	0.234	0.252	0.346	0.500	0.616	0.080	0.485		

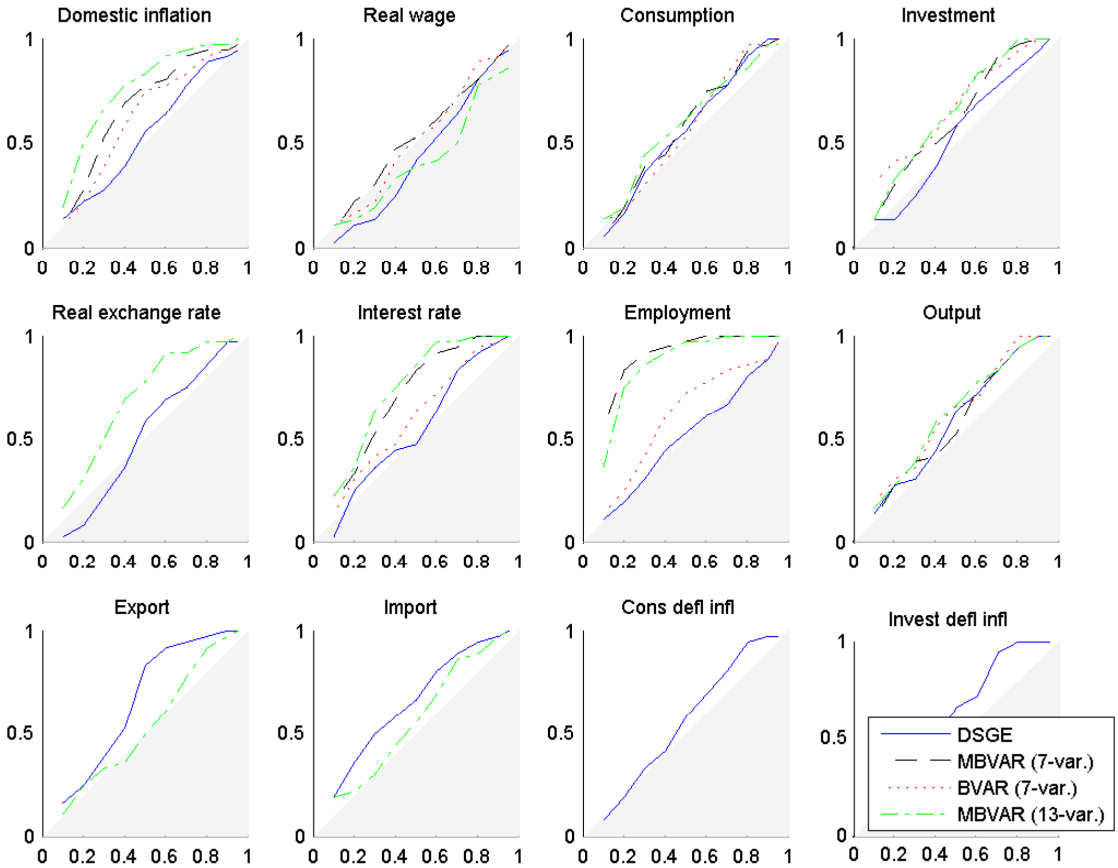

FIGURE 9. Empirical coverage probability, DSGE and BVARs, 1 quarter horizon.

In future work we plan to apply the methodology of Del Negro and Schorfheide (2004) to learn more about the misspecification of the open economy DSGE model, and test the forecasting performance of the resulting hybrid model when some of the DSGE restrictions are relaxed. However, this would still imply that a small set of variables is considered suffient to describe the economy. A broader perspective would be to estimate dynamic factor models on a much larger set of variables. For example, Stock and Watson (1999), and Giannone, Reichlin and Sala (2004) have shown that this type of models has promizing forecasting properties.

REFERENCES

[1] Adolfson, Malin, Stefan Laséen, Jesper Lindé and Mattias Villani (2005), "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through", Working Paper No. 179, Sveriges Riksbank.

[2] Altig, David, Lawrence Christiano, Martin Eichenbaum and Jesper Lindé (2003), "The Role of Monetary Policy in the Propagation of Technology Shocks", manuscript, Northwestern University.

[3] Anderson, Gary and George Moore (1985), "A Linear Algebraic Procedure for Solving Linear Perfect Foresight Models", *Economics Letters* 17(3), 247-252.

[4] Benigno, Pierpaolo (2001), "Price Stability with Imperfect Financial Integration", manuscript, New York University.

[5] Calvo, Guillermo (1983), "Staggered Prices in a Utility Maximizing Framework", *Journal of Monetary Economics* 12, 383-398.

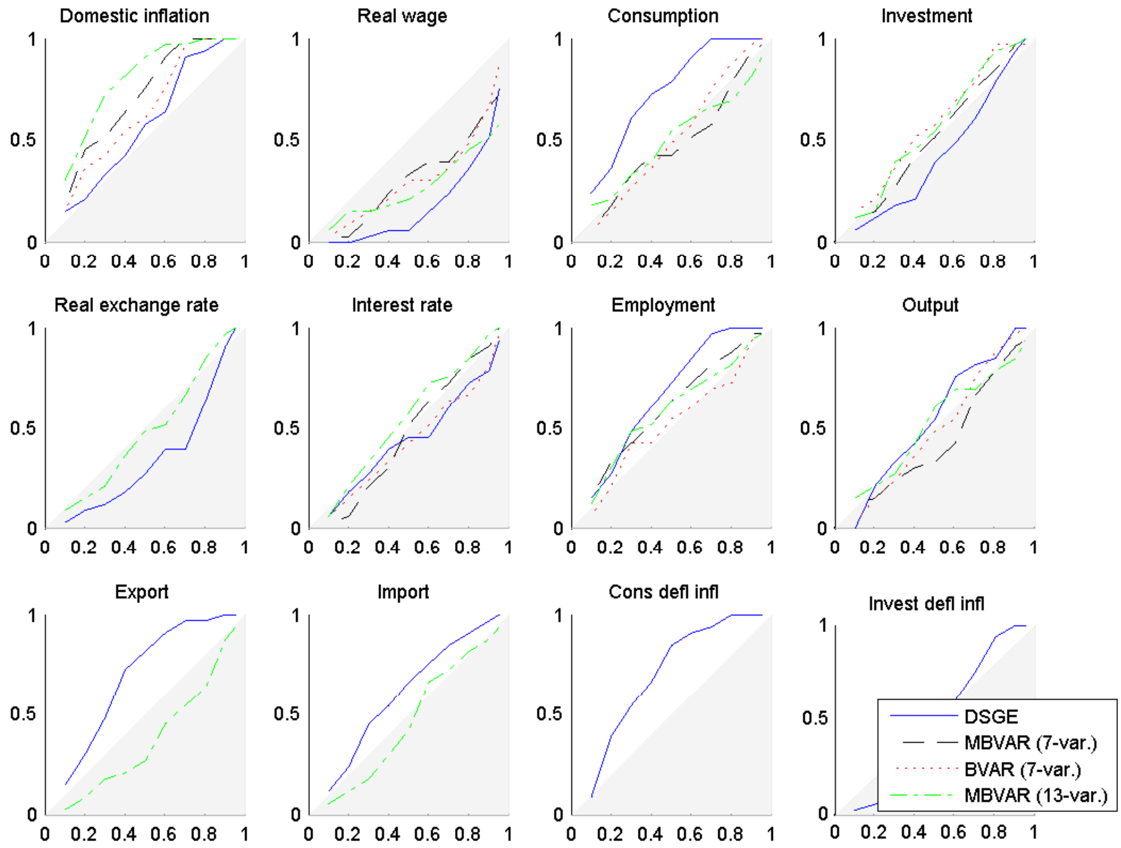

FIGURE 10. Empirical coverage probability, DSGE and BVARs, 4 quarter horizon.

[6] Christiano, Lawrence, Martin Eichenbaum and Charles Evans (2004), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy", forthcoming, *Journal of Political Economy*, 113(1), 1-45.

[7] Christoffersen, Peter (1998), "Evaluating Interval Forecasts", *International Economic Review* 39, 841-862.

[8] Clements, Michael P.and David F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

[9] Dawid, Philip (1982), "The Well-Calibrated Bayesian", *Journal of the American Statistical Association* 77, 605-610.

[10] Del Negro, Marco and Frank Schorfheide (2004), "Priors from General Equilibrium Models for VARs", *International Economic Review* 45, 643-673.

[11] Del Negro, Marco, Frank Schorfheide, Frank Smets and Raf Wouters (2004), "On the Fit and Forecasting Performance of New Keynesian Models", CEPR Discussion Paper No. 4848.

[12] Doan, Thomas (1992), RATS User's Manual, Version 4, Evanston, IL.

[13] Erceg, Christopher, Dale Henderson and Andrew Levin (2000), "Optimal Monetary Policy with Staggered Wage and Price Contracts", *Journal of Monetary Economics* 46(2), 281-313.

[14] Fagan, Gabriel, Jerome Henry and Ricardo Mestre (2001), "An Area-Wide Model (AWM) for the Euro Area", Working Paper No. 42. European Central Bank.

[15] Geweke, John (1999), "Using Simulation Methods for Bayesian Econometrics Models: Inference, Development and Communication", *Econometric Reviews* 18(1), 1-73.

[16] Giannone, Domencio, Lucrezia Reichlin and Luca Sala (2004), "Monetary Policy in Real Time", *NBER Macroeconomics Annual*, 161-200.

[17] Hamilton, James (1994), *Time Series Analysis*, Princeton University Press, Princeton.

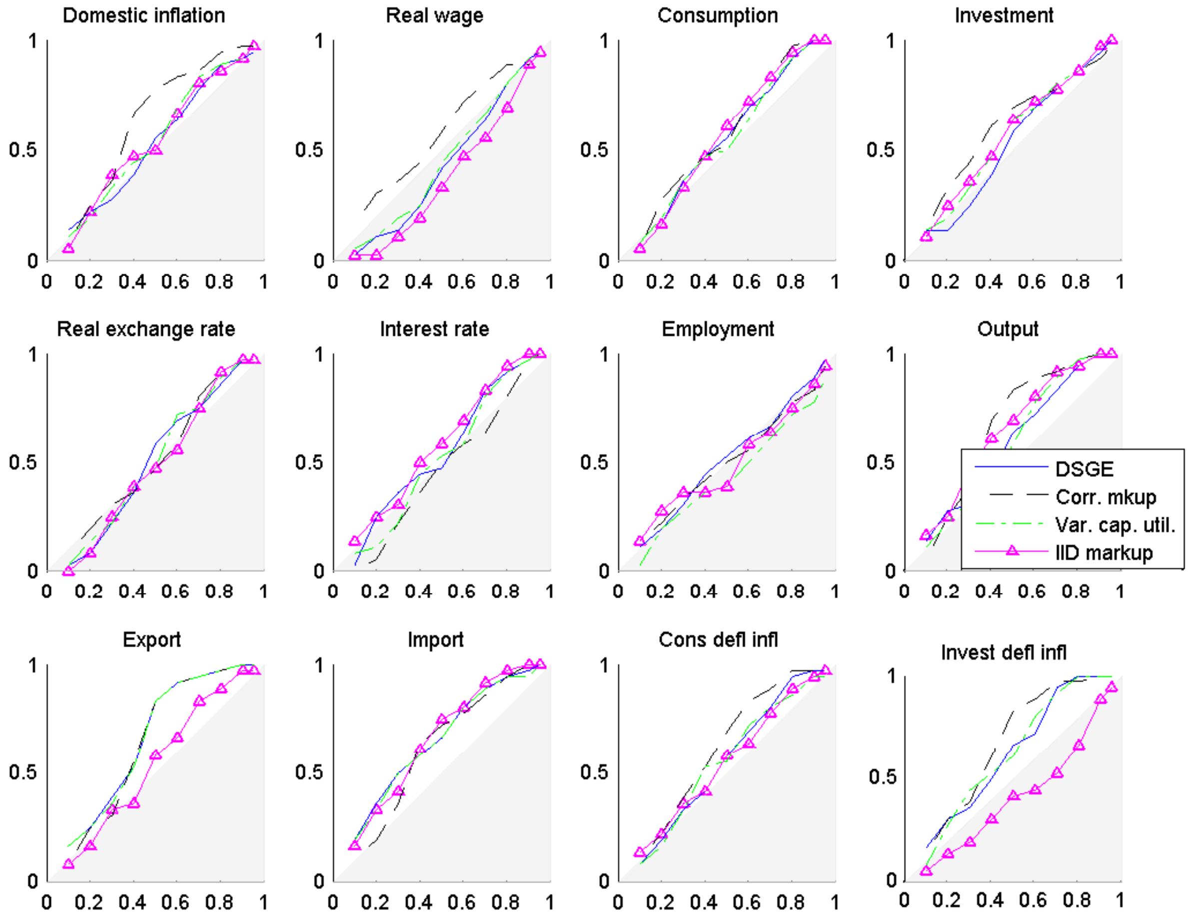

FIGURE 11. Empirical coverage probability, DSGEs, 1 quarter horizon.

[18] Jeffreys, Harold (1961), *Theory of Probability* (Third edition). Oxford: Clarendon Press.

[19] Kadiyala, Rao and Sune Karlsson (1997), "Numerical Methods for Estimation and Inference in Bayesian VAR-models", *Journal of Applied Econometrics* 12, 99-132.

[20] Litterman, Robert (1986), "Forecasting with Bayesian vector autoregressions - Five years of experience", *Journal of Business and Economic Statistics* 5, 25-38.

[21] Lundvik, Petter. (1992), "Foreign Demand and Domestic Business Cycles: Sweden 1891-1987", Chapter 3 in *Business Cycles and Growth*, Monograph Series No. 22, Institute for International Economic Studies, Stockholm University

[22] Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., Lewandowski, R., Newton, J., Parzen, E. and Winkler, R. (1982), "The accuracy of extrapolation (time series) methods: Results of a forecasting competition," *Journal of Forecasting*, 1, 111-153.

[23] Pagan, Adrian (2003), "Report on Modelling and Forecasting at the Bank of England", *Bank of England Quarterly Bulletin* 43(1), 1-29.

[24] Rotemberg, Julio and Michael Woodford (1997), "An Optimization-Based Econometric Model for the Evaluation of Monetary Policy", *NBER Macroeconomics Annual* 12, 297-346.

[25] Schorfheide, Frank (2000), "Loss Function-based Evaluation of DSGE Models", *Journal of Applied Econometrics*, 15(6), 645-670.

[26] Sims, Christopher (2003), "Comments on Smets and Wouters", unpublished notes, available at http://sims.princeton.edu/yftp/Ottawa/SWcommentSlides.pdf.

[27] Smets, Frank and Raf Wouters (2002), "Openness, Imperfect Exchange Rate Pass-Through and Monetary Policy", *Journal of Monetary Economics* 49(5), 913-940.
[28] Smets, Frank and Raf Wouters (2003), "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area", *Journal of the European Economic Association*, 1(5), 1123-1175.
[29] Smets, Frank and Raf Wouters (2004), "Forecasting with a Bayesian DSGE Model. An Application to the Euro Area", *Journal of Common Market Studies* 42(4), 841-867.
[30] Stock, James and Mark Watson (1999), "Forecasting Inflation", *Journal of Monetary Economics* 44(2), 293-335.
[31] Villani, Mattias (2005), "Inference in Vector Autoregressive Models with an Informative Prior on the Steady State", Working Paper No. 181, Sveriges Riksbank.