Forecasting Performance of an Estimated Open Economy DSGE Model for the Euro Area

Malin Adolfson
Sveriges Riksbank

Jesper Lindé
Sveriges Riksbank and IIES

Mattias Villani*
Sveriges Riksbank and Stockholm University

[Preliminary and incomplete: comments most welcome]

April 18, 2005

Abstract

This paper analyzes the forecasting performance of an estimated open economy DSGE model for the Euro area during $1994Q1 - 2002Q4$. We compare the DSGE model to various reduced form forecasting models such as VARs, BVARs, and random walks. Several univariate and multivariate measures of out-of-sample accuracy are employed. Both point forecasts and the entire forecast distribution are evaluated.

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

JEL Classification Numbers: E37; E47; C11.

* *Address*: Sveriges Riksbank, SE-103 37 Stockholm, Sweden. *E-mails*: malin.adolfson@riksbank.se, jesper.linde@riksbank.se and 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.

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 at central banks. 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 central banks, the comparison of out-of-sample forecasting properties is of particular interest, as policy actions are typically based on future inflation and output prospects. 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. The open economy aspects add complexity in terms of the number of shocks as well as the transmission of monetary policy and shocks, but they also enables us to predict "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 the difficulties to project these variables accurately. 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 more empirically 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 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.

[**Summary of the main results. Remains to be written.**]

The rest of the paper is organized as follows. Section 2 presents the key equations of the log-linearized model and reports the estimation results of four different DSGE specifications. In Section 3 we briefly discuss the alternative models used for forecasting such as vector autoregressive models and 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 estimated DSGE model

2.1. The theoretical model

This section gives an overview of the model and some key equations, and presents the log-linearized model. We refer to Adolfson, Laséen, Lindé and Villani (2005) for a more detailed description of the model.

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

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

where $C_{j,t}$, $h_{j,t}$ and $Q_{j,t}/P_t^d$ denote the j^{th} household's levels of aggregate consumption, work effort and real cash holdings, respectively. To make cash balances in (1) stationary when the economy is growing they are scaled by the unit root technology shock z_t (see below). Households consumes a basket of domestically produced goods and imported products which are supplied by domestic and importing firms, respectively. Aggregate consumption is assumed to be given by the following constant elasticity of substitution (CES) function:

$$C_t = \left[(1-\omega_c)^{1/\eta_c} \left(C_t^d\right)^{(\eta_c-1)/\eta_c} + \omega_c^{1/\eta_c} \left(C_t^m\right)^{(\eta_c-1)/\eta_c} \right]^{\eta_c/(\eta_c-1)}, \qquad (2)$$

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 can save in domestic bonds and foreign bonds, and also hold cash. Following Benigno (2001), 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. This ensures a well defined steady-state in the model.

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. 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

$$I_t = \left[(1-\omega_i)^{1/\eta_i} \left(I_t^d\right)^{(\eta_i-1)/\eta_i} + \omega_i^{1/\eta_i} \left(I_t^m\right)^{(\eta_i-1)/\eta_i} \right]^{\eta_i/(\eta_i-1)}, \qquad (3)$$

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. This gives rise to a wage equation with Calvo (1983) stickiness.

There is a continuum of intermediate domestic firms that each produce a differentiated good. These intermediate goods are sold to a retailer which transforms the intermediate products into a homogenous final good that in turn is sold to the households. The domestic firms determine the capital services and labour inputs used in their production which is exposed to unit root technology growth as in Altig et al. (2003). Production of the domestic intermediate good i follows

$$Y_{i,t} = z_t^{1-\alpha}\epsilon_t K_{i,t}^{\alpha} H_{i,t}^{1-\alpha} - z_t\phi, \tag{4}$$

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. Note also that a fixed cost is included in the production function to ensure that profits are zero in steady state.

The domestic firms, the importing and exporting firms all produce differentiated goods and set prices according to an indexation variant of the Calvo model. By including nominal rigidities in the importing and exporting sectors we allow for short-run incomplete exchange rate pass-through to both import and export prices, following for example Smets and Wouters (2002).

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.[1]

The first-order conditions of the households and the firms are log-linearized around the steady state, according to the following. The domestic (d), importing consumption (mc), importing investment (mi) and exporting (x) firms operating in this economy each have a particular Phillips curve:

$$\begin{aligned}\left(\widehat{\pi}_t^j - \widehat{\bar{\pi}}_t^c\right) &= \frac{\beta}{1+\kappa_j\beta}\left(\mathrm{E}_t\widehat{\pi}_{t+1}^j - \rho_\pi\widehat{\bar{\pi}}_t^c\right) + \frac{\kappa_j}{1+\kappa_j\beta}\left(\widehat{\pi}_{t-1}^j - \widehat{\bar{\pi}}_t^c\right) \\ &\quad - \frac{\kappa_j\beta\left(1-\rho_\pi\right)}{1+\kappa_j\beta}\widehat{\bar{\pi}}_t^c + \frac{(1-\xi_j)(1-\beta\xi_j)}{\xi_j\left(1+\kappa_j\beta\right)}\left(\widehat{mc}_t^j + \widehat{\lambda}_t^j\right),\end{aligned} \tag{5}$$

where $j = \{d, mc, mi, x\}$, $\widehat{\pi}_t^j = (\hat{P}_t^j - \hat{P}_{t-1}^j)$ denotes inflation in sector j, and $\widehat{\bar{\pi}}_t^c$ a time-varying inflation target of the central bank.[2] The ξ:s are the Calvo price stickiness parameters in each sector, and the κ:s are the indexation parameters.[3] $\widehat{\lambda}_t^d$, $\widehat{\lambda}_t^{mc}$, $\widehat{\lambda}_t^{mi}$, and $\widehat{\lambda}_t^x$ are stochastic AR(1)

[1] It should be noted that Adolfson et al. (2005) report that the fiscal shocks have small dynamic effects in the model, presumably because these shocks are transitory and do not generate any wealth effects for the infinitively lived households.

[2] A hat denotes log-linearized variables throughout the paper (i.e, $\widehat{X}_t = dX_t/X$), and variables without time-subscript steady-state values. Variables denoted with small letters have been stationarized with the unit root technology shock.

[3] For the firms that are not allowed to reoptimize their price, we adopt the indexation scheme $P_{t+1}^j = \left(\pi_t^j\right)^{\kappa_j}\left(\bar{\pi}_{t+1}^c\right)^{1-\kappa_j} P_t^j$ where $j = \{d, mc, mi, x\}$.

processes determining the time-varying markups in the four markets. The firms' marginal costs are defined as $\widehat{mc}_t^d = \alpha\left(\hat{\mu}_{z,t} + \hat{H}_t - \hat{k}_t\right) + \widehat{\bar{w}}_t + \hat{R}_t^f - \hat{\epsilon}_t$, $\widehat{mc}_t^{mc} = \widehat{P}_t^* + \widehat{S}_t - \widehat{P}_t^{mc}$, $\widehat{mc}_t^{mi} = \widehat{P}_t^* + \widehat{S}_t - \widehat{P}_t^{mi}$, and $\widehat{mc}_t^x = \widehat{P}_t^d - \widehat{S}_t - \widehat{P}_t^x$, respectively. $\hat{\mu}_{z,t}$ is the stochastic growth rate of the unit root technology shock, $\hat{H}_t$ hours worked, $\hat{k}_t$ the capital services stock, $\widehat{\bar{w}}_t$ the real wage, and $\hat{R}_t^f$ the effective nominal interest rate paid by the firms, reflecting the assumption that a fraction ν of the firms' wage bill has to be financed in advance (throughout the paper, we set $\nu = 1$). $\hat{\epsilon}_t$ is a stationary technology shock, $\widehat{P}_t^*$ the foreign price level and $\widehat{S}_t$ is the nominal exchange rate.

Under the assumption that those households that are not allowed to reoptimize their nominal wage in the current period instead update it according to the indexation scheme $W_{t+1} = (\pi_t^c)^{\kappa_w}\left(\bar{\pi}_{t+1}^c\right)^{(1-\kappa_w)}\mu_{z,t+1}W_t$, the real wage equation can be written

$$\mathrm{E}_t\left[\begin{array}{c} \eta_0\widehat{\bar{w}}_{t-1} + \eta_1\widehat{\bar{w}}_t + \eta_2\widehat{\bar{w}}_{t+1} + \eta_3\left(\hat{\pi}_t^d - \widehat{\bar{\pi}}_t^c\right) + \eta_4\left(\hat{\pi}_{t+1}^d - \rho_{\widehat{\bar{\pi}}^c}\widehat{\bar{\pi}}_t^c\right) \\ +\eta_5\left(\hat{\pi}_{t-1}^c - \widehat{\bar{\pi}}_t^c\right) + \eta_6\left(\hat{\pi}_t^c - \rho_{\widehat{\bar{\pi}}^c}\widehat{\bar{\pi}}_t^c\right) \\ +\eta_7\hat{\psi}_{z,t} + \eta_8\hat{H}_t + \eta_9\hat{\tau}_t^y + \eta_{10}\hat{\tau}_t^w + \eta_{11}\hat{\zeta}_t^h \end{array}\right] = 0, \tag{6}$$

where $\hat{\pi}_t^c$ denotes CPI inflation, $\hat{\psi}_{z,t}$ the marginal utility of one additional income unit, $\hat{\tau}_t^y$ a labour income tax, $\hat{\tau}_t^w$ a pay-roll tax assumed to be paid by the households, and $\hat{\zeta}_t^h$ a labour supply shock. The η:s are composite parameters determined by the Calvo wage stickiness ξ_w, the pay-roll tax τ^w, the labour income tax τ^y, the labour supply elasticity σ_L, the wage markup λ_w, the wage indexation κ_w, and the discount factor β.

The households' consumption preferences are subject to internal habit formation, which yields the following Euler equation for consumption expenditures:

$$\mathrm{E}_t\left[\begin{array}{c} -b\beta\mu_z\hat{c}_{t+1} + \left(\mu_z^2 + b^2\beta\right)\hat{c}_t - b\mu_z\hat{c}_{t-1} + b\mu_z\left(\hat{\mu}_{z,t} - \beta\hat{\mu}_{z,t+1}\right) + \\ +\left(\mu_z - b\beta\right)\left(\mu_z - b\right)\hat{\psi}_{z,t} + \frac{\tau^c}{1+\tau^c}\left(\mu_z - b\beta\right)\left(\mu_z - b\right)\hat{\tau}_t^c \\ +\left(\mu_z - b\beta\right)\left(\mu_z - b\right)\hat{\gamma}_t^{c,d} - \left(\mu_z - b\right)\left(\mu_z\hat{\zeta}_t^c - b\beta\hat{\zeta}_{t+1}^c\right) \end{array}\right] = 0, \tag{7}$$

where $\hat{c}_t$ is consumption, $\hat{\tau}_t^c$ a consumption tax, $\hat{\gamma}_t^{c,d}$ the relative price between consumption and domestically produced goods, $\hat{\zeta}_t^c$ a consumption preference shock, b the habit persistence parameter, and μ_z is the steady-state growth rate.

By combining the first order conditions for the domestic and foreign bond holdings we obtain the following modified uncovered interest rate parity condition:

$$\widehat{R}_t - \widehat{R}_t^* = \mathrm{E}_t\Delta\widehat{S}_{t+1} - \widetilde{\phi}_a\widehat{a}_t + \widehat{\widetilde{\phi}}_t, \tag{8}$$

where $\widehat{R}_t$ is the domestic nominal interest rate, $\widehat{R}_t^*$ the foreign nominal interest rate, $\widehat{a}_t$ the net foreign asset position, and $\widehat{\widetilde{\phi}}_t$ a shock to the risk premium. Because of our assumption of imperfect integration in the international financial markets, the net foreign asset position enters.

The households' first order conditions for the physical capital stock $(\widehat{\bar{k}}_t)$, investment $(\hat{\imath}_t)$, and the utilization rate ($\widehat{u}_t = \widehat{k}_t - \widehat{\bar{k}}_t$, where $\widehat{k}_t$ denotes capital services) can be written:

$$\begin{array}{r} \hat{\psi}_{z,t} + \mathrm{E}_t\hat{\mu}_{z,t+1} - \mathrm{E}_t\hat{\psi}_{z,t+1} - \frac{\beta(1-\delta)}{\mu_z}\mathrm{E}_t\hat{P}_{k',t+1} + \hat{P}_{k',t} \\ -\frac{\mu_z-\beta(1-\delta)}{\mu_z}\mathrm{E}_t\hat{r}_{t+1}^k + \frac{\tau^k}{(1-\tau^k)}\frac{\mu_z-\beta(1-\delta)}{\mu_z}\mathrm{E}_t\hat{\tau}_{t+1}^k = 0, \end{array} \tag{9}$$

$$\hat{P}_{k',t} + \hat{\Upsilon}_t - \hat{\gamma}_t^{i,d} - \mu_z^2\tilde{S}''\left[\left(\hat{\imath}_t - \hat{\imath}_{t-1}\right) - \beta\left(\hat{\imath}_{t+1} - \hat{\imath}_t\right) + \hat{\mu}_{z,t} - \beta\mathrm{E}_t\hat{\mu}_{z,t+1}\right] = 0, \tag{10}$$

$$\widehat{u}_t = \frac{1}{\sigma_a}\widehat{r}_t^k - \frac{1}{\sigma_a}\frac{\tau^k}{(1-\tau^k)}\widehat{\tau}_t^k, \tag{11}$$

where $\hat{P}_{k',t}$ is the price of capital, $\widehat{r}_t^k$ the firms' real rental rate of capital services given by $\hat{r}_t^k = \hat{\mu}_{z,t} + \widehat{\bar{w}}_t + \hat{R}_t^f + \hat{H}_t - \hat{k}_t$, $\hat{\Upsilon}_t$ an investment specific technology shock, $\hat{\gamma}_t^{i,d}$ the relative price between investment and domestically produced goods, $\hat{\tau}_t^k$ a capital income tax, $\tilde{S}''$ the adjustment cost of changing investments, δ the depreciation rate, and σ_a the cost of varying the capital utilization rate.

The log-linearized law of motion for the physical capital stock is given by

$$\widehat{\bar{k}}_{t+1} = (1-\delta)\frac{1}{\mu_z}\widehat{\bar{k}}_t - (1-\delta)\frac{1}{\mu_z}\hat{\mu}_{z,t} + \left(1-(1-\delta)\frac{1}{\mu_z}\right)\hat{\Upsilon}_t + \left(1-(1-\delta)\frac{1}{\mu_z}\right)\hat{\imath}_t. \tag{12}$$

The evolution of net foreign assets at the aggregate level satisfies

$$\begin{aligned}\hat{a}_t \;=\;& -y^*\widehat{mc}_t^x - \eta_f y_t^* \widehat{\gamma}_t^{x,*} + y^*\hat{y}_t^* + y^*\widehat{\tilde{z}}_t^* + (c^m + i^m)\,\widehat{\gamma}_t^f \\ & -c^m\left[-\eta_c(1-\omega_c)\left(\gamma^{c,d}\right)^{-(1-\eta_c)}\widehat{\gamma}_t^{mc,d} + \widehat{c}_t\right] \\ & -i^m\left[-\eta_i(1-\omega_i)\left(\gamma^{i,d}\right)^{-(1-\eta_i)}\widehat{\gamma}_t^{mi,d} + \widehat{\imath}_t\right] + \frac{R}{\pi\mu_z}\hat{a}_{t-1},\end{aligned} \tag{13}$$

where $\hat{y}_t^*$ denotes foreign output, $\widehat{\tilde{z}}_t^*$ is a stationary shock which measures the degree of asymmetry in the technological progress in the domestic economy versus the rest of the world, and $\widehat{\gamma}_t^{x,*}$,$\widehat{\gamma}_t^f$, $\widehat{\gamma}_t^{mc,d}$ and $\widehat{\gamma}_t^{mi,d}$ are relative prices defined as $\widehat{\gamma}_t^{x,*} = \hat{P}_t^x - \hat{P}_t^*$, $\widehat{\gamma}_t^f = \hat{P}_t^d - \hat{S}_t - \hat{P}_t^*$, $\widehat{\gamma}_t^{mc,d} = \hat{P}_t^{mc} - \hat{P}_t^d$ and $\widehat{\gamma}_t^{mi,d} = \hat{P}_t^{mi} - \hat{P}_t^d$, respectively.

The log-linearized first order conditions for money balances and the households cash holdings are, respectively:

$$\mathrm{E}_t\left[-\mu\hat{\psi}_{z,t} + \mu\hat{\psi}_{z,t+1} - \mu\hat{\mu}_{z,t+1} + \left(\mu - \beta\tau^k\right)\hat{R}_t - \mu\hat{\pi}_{t+1} + \frac{\tau^k}{1-\tau^k}(\beta-\mu)\,\hat{\tau}_{t+1}^k\right] = 0, \tag{14}$$

$$q_t = \frac{1}{\sigma_q}\left[\frac{\tau^k}{1-\tau^k}\widehat{\tau}_t^k - \widehat{\psi}_{z,t} - \frac{R}{R-1}\widehat{R}_{t-1}\right]. \tag{15}$$

The log-linearized aggregate resource constraint is

$$\begin{aligned}& (1-\omega_c)\left(\gamma^{c,d}\right)^{\eta_c}\frac{c}{y}\left(\hat{c}_t + \eta_c\widehat{\gamma}_t^{c,d}\right) + (1-\omega_i)\left(\gamma^{i,d}\right)^{\eta_i}\frac{i}{y}\left(\hat{\imath}_t + \eta_i\widehat{\gamma}_t^{i,d}\right) \\ & +\frac{g}{y}\hat{g}_t + \frac{y^*}{y}\left(\hat{y}_t^* - \eta_f\widehat{\gamma}_t^{x,*} + \widehat{\tilde{z}}_t^*\right) \\ =\; & \lambda_f\left(\hat{\epsilon}_t + \alpha\left(\hat{k}_t - \hat{\mu}_{z,t}\right) + (1-\alpha)\hat{H}_t\right) - \left(1-\tau^k\right)r^k\frac{\bar{k}}{y}\frac{1}{\mu_z}\left(\hat{k}_t - \widehat{\bar{k}}_t\right),\end{aligned} \tag{16}$$

where $\widehat{\gamma}_t^{i,d}$ is the relative price between investment and domestically produced goods.

To clear the loan market, the demand for liquidity from the firms (which are financing their wage bills) must equal the supplied deposits of the households plus the monetary injection by the central bank:

$$\nu\bar{w}H\left(\hat{\nu}_t + \widehat{\bar{w}}_t + \hat{H}_t\right) = \frac{\mu\bar{m}}{\pi\mu_z}\left(\hat{\mu}_t + \widehat{\bar{m}}_t - \hat{\pi}_t - \hat{\mu}_{z,t}\right) - q\hat{q}_t, \tag{17}$$

Following Smets and Wouters (2003), monetary policy is approximated with the instrument rule

$$\begin{aligned} \widehat{R}_t &= \rho_R \widehat{R}_{t-1} + (1-\rho_R)\left[\widehat{\bar{\pi}}_t^c + r_\pi\left(\hat{\pi}_{t-1}^c - \widehat{\bar{\pi}}_t^c\right) + r_y \hat{y}_{t-1} + r_x \hat{x}_{t-1}\right] \\ &\quad + r_{\Delta\pi}\left(\hat{\pi}_t^c - \hat{\pi}_{t-1}^c\right) + r_{\Delta y}\Delta\hat{y}_t + \varepsilon_{R,t}, \end{aligned} \tag{18}$$

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 $\left(\hat{\pi}_t^c - \widehat{\bar{\pi}}_t^c\right)$, 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.

2.2. Estimation

To estimate the model we use quarterly euro area data for the period 1970Q1-2002Q4. The data set employed here was first constructed by Fagan et al. (2001).[4] We include a large set of variables when we estimate the model in order to facilitate identification of the parameters, and match the following 15 variables: the domestic inflation rate π_t; the short-run interest rate R_t; employment $\widehat{E}_t$; the growth rates in consumption Δc_t(Δ denotes the first difference operator), investment Δi_t, GDP Δy_t, the real wage $\Delta \overline{w}_t$, exports $\Delta \widetilde{X}_t$, imports $\Delta \widetilde{M}_t$, the consumption deflator $\pi_t^{def,c}$ and the investment deflator $\pi_t^{def,i}$; the real exchange rate x_t; foreign inflation π_t^*; the foreign interest rate R_t^*; and the growth rate in foreign output Δy_t^*.[5] The reason for modeling the real variables in growth rates is that the unit root technology shock induces a stochastic trend in the levels of these variables. To calculate the likelihood function of the observed variables we apply the Kalman filter.[6] In Figure 1 the actual data series are depicted. Note that annualized rates of the quarterly inflation and interest rates series are shown, while employment and the real exchange rate are measured as percentage deviations around the mean.

A number of parameters are kept fixed throughout the estimation procedure. Most of these parameters can be related to the steady-state values of the observed variables in the model, and are therefore calibrated so as to match the sample mean of these. Table 1 reports the calibrated parameters along with the implied steady state values of some key variables.

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. Second, draws from the joint posterior are generated using the Metropolis-Hastings algorithm (see Schorfheide (2000) for details). In Table 2 we report the posterior mode.

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

[4] 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 R_t^*.

[5] There is no (official) data on aggregate hours worked, $\hat{H}_t$, available for the euro area. Therefore, we use employment $\hat{E}_t$ 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): $\Delta\hat{E}_t = \beta E_t \Delta\hat{E}_{t+1} + \frac{(1-\xi_e)(1-\beta\xi_e)}{\xi_e}\left(\hat{H}_t - \hat{E}_t\right)$. 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.

[6] We use the period 1970Q1-1979Q4 to form a prior on the unobserved state variables in 1979Q4, and then use the period 1980Q1-2002Q4 for inference.

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 markups 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.

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. The first is a "closed economy" specification composed of the seven domestic variables π_t, R_t, $\widehat{E}_t$, Δc_t, Δi_t, Δy_t, $\Delta \overline{w}_t$, and the second is an "open economy" specification which additionally includes the variables $\Delta \widetilde{X}_t$, $\Delta \widetilde{M}_t$, x_t, π_t^*, R_t^*, Δy_t^*. Note that the consumption and investment deflators, $\pi_t^{def,c}$and $\pi_t^{def,i}$, have been excluded in the VARs for reasons of parsimony. The usual parametrization of the VAR models reads

$$\Pi(L)x_t = \Phi d_t + \varepsilon_t, \tag{19}$$

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

$$\Pi(L)(x_t - \Psi d_t) = \varepsilon_t. \tag{20}$$

This somewhat non-standard parametrization of the VAR model in (20) 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 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 (19) and (20) 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 default values on the hyperparameters in the priors according to the following: 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. 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.

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 (19) and the Gibbs sampler in Villani (2005) for the specification in (20).

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. (19) and eq. (20), respectively. In addition to this we have also estimated the 7- and 13-variables system with maximum likelihood. To save space we have chosen 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 our DSGE model 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 coverage rates of 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 this 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 $1, 2, 4$, and 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 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 considerations 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 micro 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 forecast

Let $\hat{x}_{t+h|t}$ denote the h-step-ahead posterior median forecast of x_{t+h}, standing at time t, and define $e_t(h) = x_{t+h} - \hat{x}_{t+h|t}$ is 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):

$$MAE_i(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} |e_{i,t}(h)|, \tag{21}$$

$$RMSE_i(h) = \sqrt{N_h^{-1} \sum_{t=T}^{T+N_h-1} e_{i,t}^2(h)}, \tag{22}$$

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. The predictive distribution includes uncertainty regarding both shocks and measurement errors.

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

$$\Omega_M(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} \tilde{e}_t(h)\tilde{e}_t'(h). \tag{23}$$

where $\tilde{e}_t(h) = M^{-1/2}e_t(h)$. The matrix M, the sample covariance matrix of the time series based on data from $1993Q1 - 2002Q4$, acts as a scaling matrix that accounts for the fact that the time series may be more or less intrinsically predictable in absolute terms. For example, we expect the models to better project variables such as consumption and output, while investment, imports, exports and in particular the real exchange rate are highly volatile series that are more difficult to predict. In order for these series not to entirely dominate the trace statistic defined below, the forecast errors are scaled. 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.

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. The information that one model is outperforming another in terms of e.g. $\ln|\Omega|$ naturally raises the question: in which dimensions is the gain in forecasting accuracy most substantial? To answer this question we perform a singular value decomposition of Ω: $\Omega = V\Lambda V'$,where $V = (v_1, ..., v_k)$, $V'V = I_k$, is the matrix of eigenvectors, $\Lambda = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_k)$ 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 $\text{tr}(\Omega) = \sum_{i=1}^{k} \lambda_i$, it is clear that $\text{tr}(\Omega)$ 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 $\text{tr}(\Omega) \to \sum_{i=1}^{k-1} \lambda_i$, but $\ln|\Omega| \to -\infty$, for any values of λ_i, $i = 1, ..., k-1$. A variable with a large (squared) coeffcient in the first principal component is thus a major contributor to $\text{tr}(\Omega)$, 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. If e.g. v_{11}^2 is close to unity (note that $\sum_{j=1}^{k} v_{j1}^2 = 1$, since eigenvectors have unit length), then the first eigenvalue measures essentially the MSE of the first variable, which in turn dominates $\text{tr}(\Omega)$, at least if λ_1 is large in comparison to the other eigenvalues.

4.4. Measuring the accuracy of density forecasts

The measures of forecasting performance presented so far values 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 steps-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-steps 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

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

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.[7] 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

[7] 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.

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 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

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

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, $\hat{\alpha}_h = N_h^{-1} \sum_{n=1}^{N_h} I_{T+1}^\alpha(h)$, 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

$$\begin{aligned} H_0 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \overset{iid}{\sim} Bern(\alpha) \\ H_1 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \overset{iid}{\sim} Bern(\pi) \\ H_2 &: \{I_t^\alpha(1)\}_{t=T+1}^{T+N_1} \sim Markov(\pi_{01}, \pi_{11}), \end{aligned} \tag{24}$$

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 a correct 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 2a shows the root mean squared forecast errors in yearly percentage terms at the $1, 2, 4$, and 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 four most recent data observations). The mean absolute forecast errors give similar results and to save space we have chosen only to report the RMSEs here. We can 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. At shorter horizons (1 and 2 quarters) the DSGE model also seems to project domestic inflation, consumption, employment, and the consumption deflator inflation very well. For output the DSGE model does slightly better forecasts than the BVARs at shorter horizons but looses somewhat in the medium run. Note also that the one- and two-step-ahead forecast from the DSGE model beats 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 classical VAR model on most variables and horizons (see Figures 2a and 2c). 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 BVARs stay around 0.7%. This is somewhat surprising given that we believe the DSGE model's theoretical structure to matter more in the long run, and therefore to have an advantage over the BVARs in the forecasting performance at those particular horizons.

To understand this, Figure 2b depicts the RMSEs for the four different specifications of the DSGE model. 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 BVAR evidence). The baseline model seems to overpredict both inflation and the real wage more often at longer horizons than the model with correlated markup shocks (not shown). One 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. 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). 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.

Table 4 shows the multivariate accuracy measures for the point forecasts, the log determinant and the trace of the MSSE matrix (see equation (23)). In order to be able to compare the multivariate measures across the different models, we have chosen only to include the variables that are common to all models. The matrices are therefore based on the forecast errors for 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 to have better accuracy on the one and two 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.

A closer look at the multivariate measures using the spectral decomposition of the MSSE matrix discussed in Section 4.3 explains the seemingly incompatible results in Figures 2a-c, on the one hand, and the multivariate measures in Table 4 results, on the other. Figure 4 displays the eigenvalues of the MSSE matrix, both on original and log scale, at the $1, 2, 4$ and 8 quarter horizon for four of the models. The log determinant statistic equals the sum of the log eigenvalues of the MSSE matrix. It is therefore clear from Figure 4 that the large difference in forecasting performance between the DGSEs and BVARs captured by this statistic at the first and second quarter horizon is dominated by the smallest eigenvalue. The variables which account for the major part of the last principal component are therefore responsible for the DSGE models inferior forecast performance at the short horizons. Looking at the subgraphs in the right column of Figure 5, which depicts the relative weight of the variables in the eigenvector with smallest eigenvalue (v_{jk}^2 for the jth variable), it is clear that this principal component is essentially the forecast errors of the employment series. This result holds for all horizons and four models. The short run RMSEs of the employment series in Figure 2a for the different models do not appear that large, but the *relative* difference between the DSGE models and the BVARs are substantial: the RMSE of employment 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 employment is still the driving force of the smallest eigenvalues at horizons 4 and 8, but here the difference in log determinant statistic across models is no longer dominated by this eigenvalue (see Figure 4). At the four quarter horizon it is actually the largest eigenvalue which is dominating the comparison. The determinants of this eigenvector are given in the left hand column of Figure 5. 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 predicted more accurately than the benchmark DSGE. Finally, on the eight quarter horizon the picture is more complicated, but the differences in log determinant are mostly explained by the maximal and the minimal eigenvalues. The benchmark DSGE scores well on the smallest eigenvalue (good employment forecasts), but is worse than the other models on the maximal eigenvalue (large RMSEs for inflation and wage forecasts). The DSGE with correlated markup shocks and the seven-variable BVAR show essentially the opposite results.

The eigenvalues on the original scale in Figure 4 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 5 may similarly be used to investigate the differences in the trace statistic.

In Figure 2c we display the RMSEs for the various VAR systems. We see that the 7 and 13-variables models have about the same point forecast accuracy for the domestic variables. The MBVARs with a prior on the steady state seems to do slightly better in terms of the forecasts on inflation, but the differences are rather small. On the other hand, the MBVAR models seem instead to perform worse on some of the real variables such as the real wage and consumption. Turning to the multivariate measures we find that the MBVAR does a good job in capturing the joint accuracy of the forecasted variables.

5.2. Density forecasts

Figure 3 shows the accuracy of the 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 in Figure 3. 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 exact numbers in Figure 3 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 the actual coverage is further away from the intended forecast density. A reason for the worse properties of the DSGE models could be the internal propagation of the disturbances hitting the economy. The processes for disturbances such as the asymmetric technology shock and the risk premium shock are very near the unit-root, which implies that the effects of these shocks can amplify over the horizon and generate wider uncertainty bands.

Table 4 reports the log predictive density score. Once again we have chosen to include only the subset of variables that are mutual to the different models we evaluate, i.e., π_t, $\hat{w}_t$, $\hat{c}_t$, $\hat{\imath}_t$, $\hat{E}_t$, R_t and $\hat{y}_t$. The LPDS appear to suggest a better overall forecast density for the BVARs, at least at the shorter horizons, while the DSGE model gain ground on the longest horizon. Note also that, even if the differences are small, the BVAR without a dummy variable (capturing the shift in nominal data) actually performs best in terms of the LPDS, except at the 4 quarter horizon. As seen above the opposite holds true when examining the point forecasts. In that case the priors on the steady state do seem to matter for the forecasting performance at longer horizons.

The joint hypothesis test on independence and coverage of the forecast interval is presented in Table 5. The table shows the posterior probabilities of the three models in equation (24) for the 75% forecast interval. A well-calibrated forecasting model should place more probability mass on H_0 where the forecast errors are independent together with the right empirical coverage for the forecast interval. From the table follows that the baseline DSGE model has most probability mass on the preferred hypothesis 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 3 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 somewhat too large.

5.3. Marginal likelihood and out-of-sample performance

From Table 2 we see that the marginal likelihood gives overwhelming emphasis to the baseline DSGE model with no variable capital utilization. The posterior odds for the baseline model and the models with either variable capital utilization, persistent markup shocks or iid markup shocks are 0.9976, 0.0003, 0.002 and 0, respectively. As disussed in Section 4.2, the theoretical correspondence between the marginal likelihood of a model and its log preditive density scores (LPDS) is rather weak, even for the one-step-ahead forecasts. Comparing the marginal likelihoods of the DSGE models to the LPDS in Table 4, we nevertheless see a remarkable correspondence in the ranking of DSGE models.

Turning to the other measures of forecasting accuracy presented in the paper, the use of the marginal likelihood as a summary of out-of-sample forecasting performance is highly questionable. At short horizons, all four models generate more or less the same RMSE for variables such as domestic inflation, output, the real exchange rate, exports and imports (see Figure 2b). 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 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 odds in terms of the point forecast accuracy.

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. Note lastly that the DSGE model outperforms the classical VAR model in terms of the RMSEs on all variables and horizons.

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 longer horizons (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 capture the multivariate forecast density somewhat better at longer horizons, while the BVARs have an overall forecasting advantage predominately at shorter horizons.

References

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.

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

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

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

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.

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

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

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

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.

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

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.

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

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

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

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

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

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

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.

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

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

Schorfheide, Frank (2000), “Loss Function-based Evaluation of DSGE Models”, *Journal of Applied Econometrics*, 15(6), 645-670.

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

Smets, Frank and Raf Wouters (2002), “Openness, Imperfect Exchange Rate Pass-Through and Monetary Policy”, *Journal of Monetary Economics* 49(5), 913-940.

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.

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.

Villani, Mattias (2005), ”Inference in Vector Autoregressive Models with an Informative Prior on the Steady State”, Working Paper No. 181, Sveriges Riksbank.

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
$S_{t+1} = S_t$	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 steady state values given by these parameters evaluated at the prior mode.

Table 3: 95% prior probability intervals of Ψ

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

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

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.

Table 2: Prior and posterior distributions

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

Table 4: Multivariate accuracy measures

Horizon	Model	Log determinant statistic	Trace statistic	Log predictive density score
1Q	DSGE, Benchmark	-14.387	2.248	*7.126*
	DSGE, with variable capital utilization	-14.133	2.353	8.165
	DSGE, correlated markup shocks	-14.508	2.196	7.517
	DSGE, all markup shocks iid	-14.035	2.536	8.424
	7-variables BVAR (mean adjusted)	<u>-16.540</u>	**2.036**	8.236
	7-variables BVAR (standard)	*-16.491*	<u>2.040</u>	<u>5.452</u>
	7-variables BVAR (no break/standard)	**-16.710**	*2.100*	**5.102**
	13-variables BVAR (mean adjusted)	-16.116	2.490	9.050
2Q	DSGE, Benchmark	-8.041	6.542	14.387
	DSGE, with variable capital utilization	-7.905	6.778	15.320
	DSGE, correlated markup shocks	-8.111	5.451	14.882
	DSGE, all markup shocks iid	-7.634	7.654	16.014
	7-variables BVAR (mean adjusted)	**-9.679**	**4.793**	14.109
	7-variables BVAR (standard)	*-9.396*	<u>5.051</u>	<u>13.052</u>
	7-variables BVAR (no break/standard)	<u>-9.664</u>	*5.203*	**12.382**
	13-variables BVAR (mean adjusted)	-9.098	6.449	15.529
4Q	DSGE, Benchmark	-1.607	22.904	22.929
	DSGE, with variable capital utilization	-1.484	21.949	23.979
	DSGE, correlated markup shocks	<u>-2.049</u>	<u>13.064</u>	23.089
	DSGE, all markup shocks iid	-1.219	25.468	25.027
	7-variables BVAR (mean adjusted)	**-2.490**	**12.774**	<u>21.880</u>
	7-variables BVAR (standard)	-1.315	*14.831*	*22.603*
	7-variables BVAR (no break/standard)	*-1.863*	15.116	**21.351**
	13-variables BVAR (mean adjusted)	-1.286	21.945	24.708
8Q	DSGE, Benchmark	<u>-1.077</u>	33.696	<u>26.652</u>
	DSGE, with variable capital utilization	*-0.742*	23.412	27.334
	DSGE, correlated markup shocks	**-1.468**	**12.621**	27.708
	DSGE, all markup shocks iid	-0.715	30.323	27.364
	7-variables BVAR (mean adjusted)	-0.419	<u>16.458</u>	26.920
	7-variables BVAR (standard)	1.505	23.187	*26.817*
	7-variables BVAR (no break/standard)	0.942	24.360	**26.580**
	13-variables BVAR (mean adjusted)	0.705	*22.743*	30.364

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

Table 5: Calibration inference for forecast intervals with a coverage probability of 75%

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	**0.710**	**0.768**	**0.444**	**0.594**	**0.771**	**0.444**	**0.719**	0.373	0.067	0.418	0.214	0.012
		H_1	0.209	0.162	0.218	0.292	0.164	0.218	0.131	**0.379**	**0.667**	**0.424**	**0.578**	**0.646**
		H_2	0.082	0.070	0.337	0.114	0.065	0.337	0.149	0.248	0.265	0.159	0.208	0.342
	DSGE, with variable capital utilization	H_0	**0.583**	**0.719**	0.117	**0.594**	**0.583**	**0.444**	**0.585**	0.214	0.067	0.418	**0.586**	0.067
		H_1	0.286	0.200	0.057	0.292	0.286	0.218	0.163	**0.578**	**0.667**	**0.424**	0.287	**0.667**
		H_2	0.131	0.081	**0.826**	0.114	0.131	0.337	0.252	0.208	0.265	0.159	0.127	0.265
	DSGE, correlated markup shocks	H_0	0.144	**0.583**	0.024	**0.503**	0.373	**0.663**	0.053	0.067	0.067	0.067	0.012	0.012
		H_1	0.389	0.286	0.233	0.247	**0.379**	0.121	0.011	**0.667**	**0.667**	**0.667**	**0.646**	**0.646**
		H_2	**0.467**	0.131	**0.743**	0.250	0.248	0.216	**0.936**	0.265	0.265	0.265	0.342	0.342
	DSGE, all markup shocks iid	H_0	**0.710**	**0.499**	0.159	**0.594**	**0.583**	0.067	**0.794**	0.067	0.373	0.067	**0.583**	0.083
		H_1	0.209	0.353	0.161	0.292	0.286	**0.667**	0.145	**0.667**	**0.379**	**0.667**	0.286	**0.580**
		H_2	0.082	0.149	**0.680**	0.114	0.131	0.265	0.061	0.265	0.248	0.265	0.131	0.337
	7-variables BVAR	H_0	0.144	**0.656**	**0.586**	0.214		0.012	0.001	0.019				
	(mean adjusted)	H_1	0.389	0.140	0.287	**0.578**		**0.646**	0.500	0.050				
		H_2	**0.467**	0.204	0.127	0.208		0.342	0.500	**0.931**				
	7-variables BVAR	H_0	0.373	**0.713**	0.373	0.036		0.214	**0.710**	0.013				
	(standard)	H_1	**0.379**	0.210	**0.379**	0.358		**0.578**	0.209	**0.718**				
		H_2	0.248	0.077	0.248	**0.606**		0.208	0.082	0.269				
	7-variables BVAR	H_0	0.373	**0.719**	0.214	0.067		0.012	0.001	0.067				
	(no break/mean adj.)	H_1	**0.379**	0.131	**0.578**	**0.667**		**0.646**	0.500	**0.667**				
		H_2	0.248	0.149	0.208	0.265		0.342	0.500	0.265				
	7-variables BVAR	H_0	0.373	**0.656**	0.418	0.012		0.418	**0.710**	0.001				
	(no break/standard)	H_1	**0.379**	0.140	**0.424**	**0.646**		**0.424**	0.209	0.500				
		H_2	0.248	0.204	0.159	0.342		0.159	0.082	0.500				
	13-variables BVAR	H_0	0.024	**0.503**	**0.444**	0.012	0.012	0.001	0.001	0.214	**0.583**	0.159		
	(mean adjusted)	H_1	0.233	0.140	0.218	**0.646**	**0.646**	0.500	0.500	**0.578**	0.286	0.161		
		H_2	**0.743**	0.357	0.337	0.342	0.342	0.500	0.500	0.208	0.131	**0.680**		

Figure 1: Actual data $1980Q1 - 2002Q4$

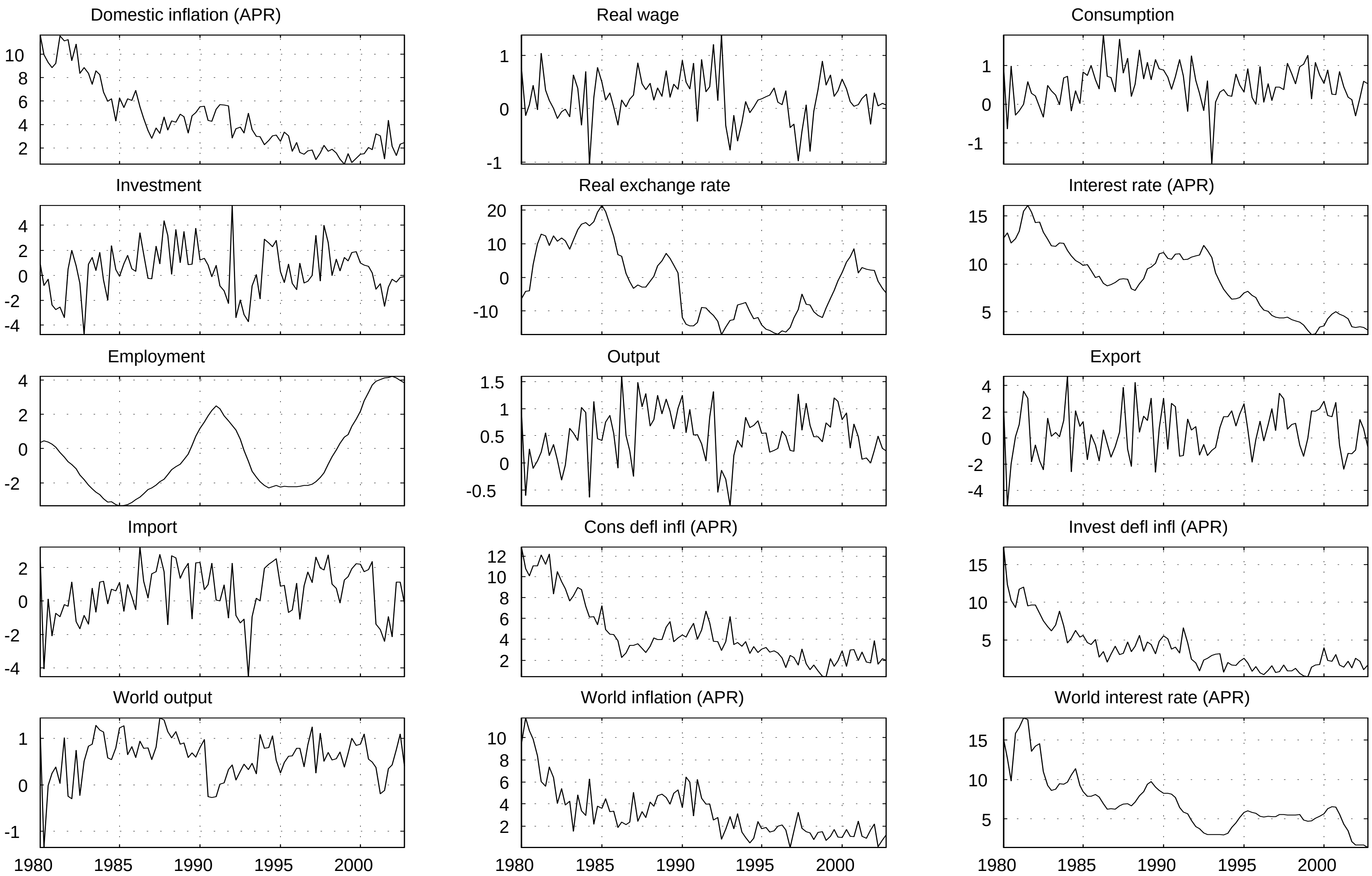


Figure 2a: Root mean squared forecast errors, DSGE, BVARs, MLVAR, and naive

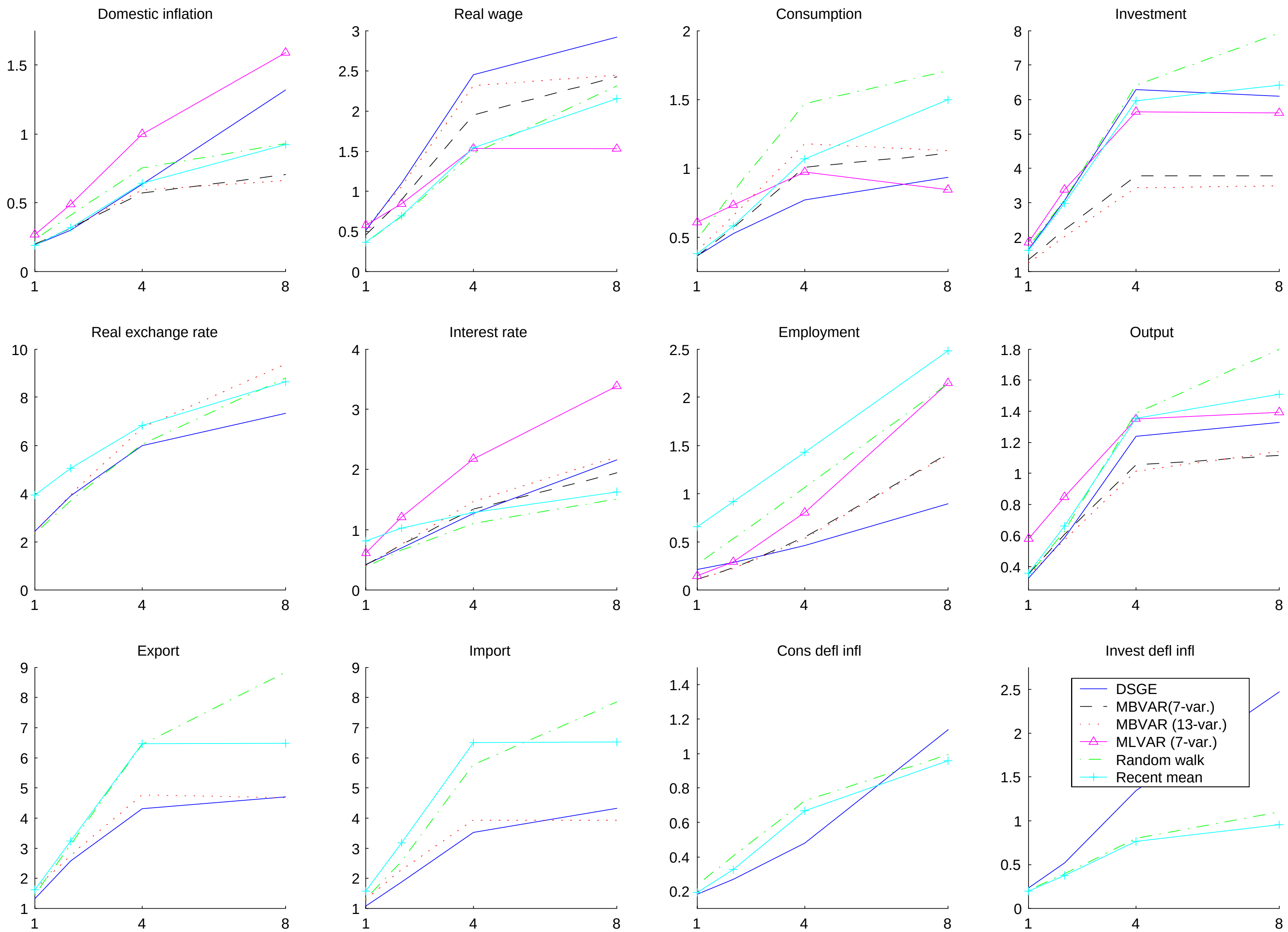


Figure 2b: Root mean squared forecast errors, DSGE models

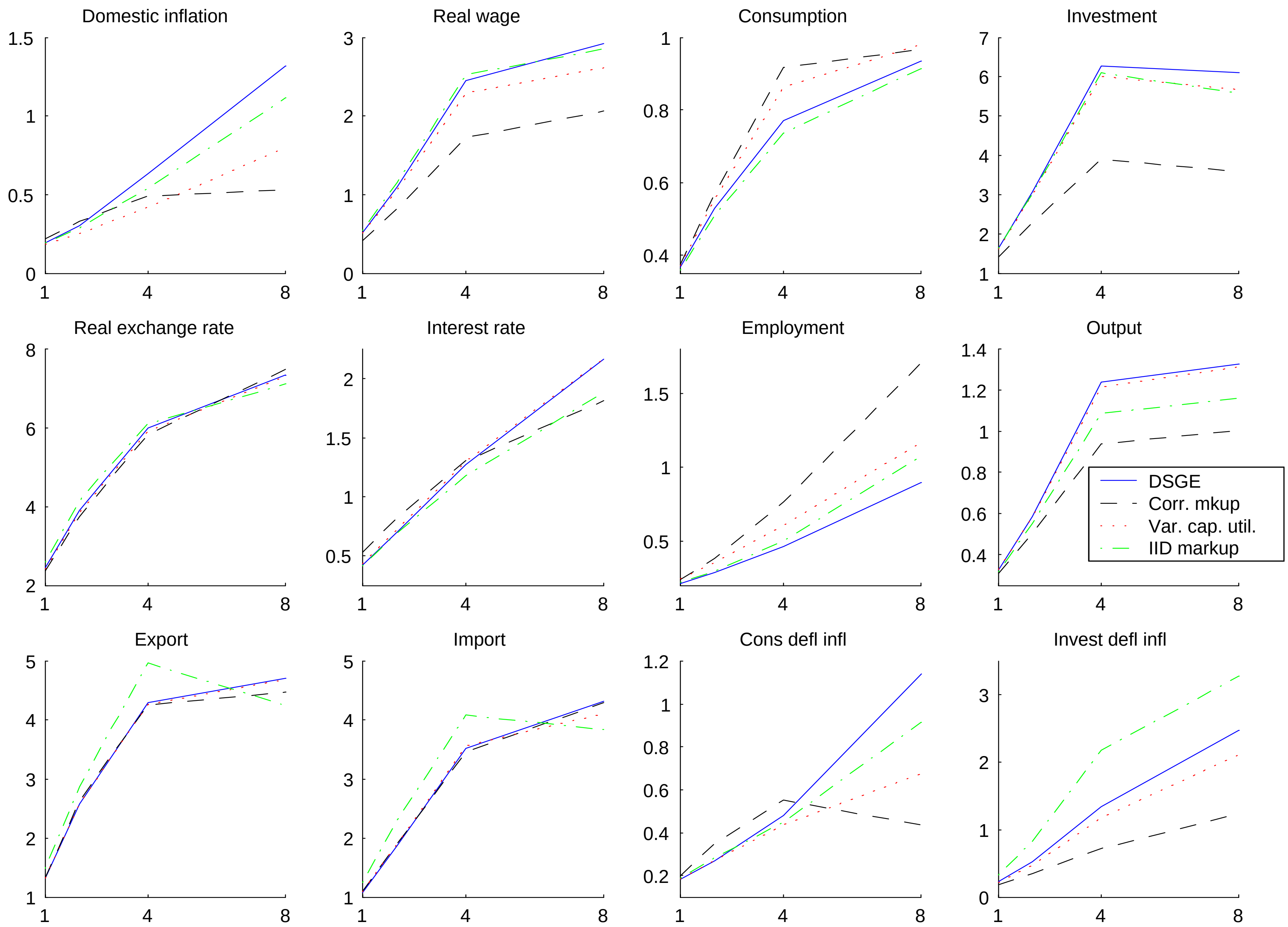


Figure 2c: Root mean squared forecast errors, VAR models

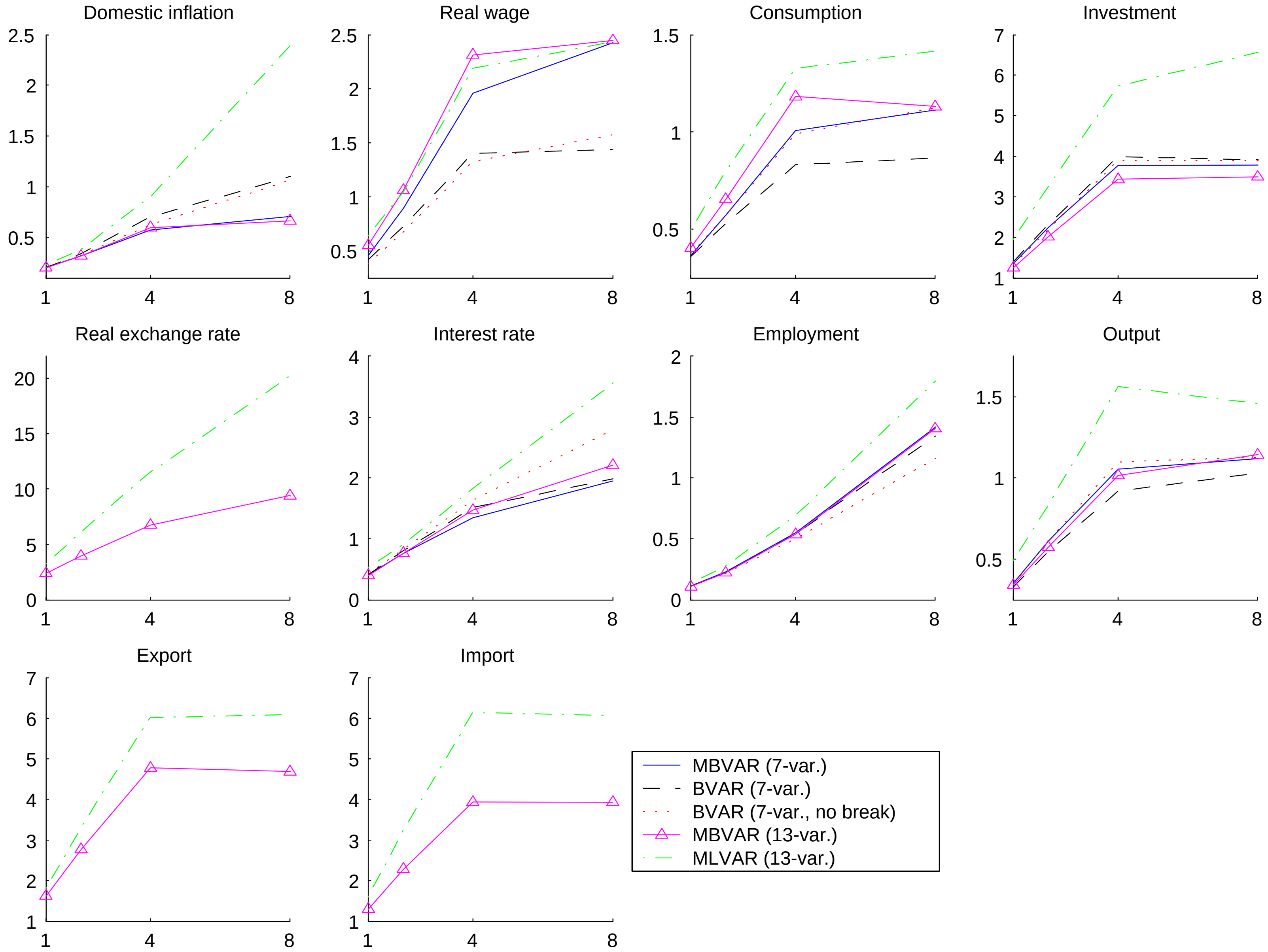


Figure 3a: Empirical coverage probability, DSGE and BVARs 1 quarter horizon

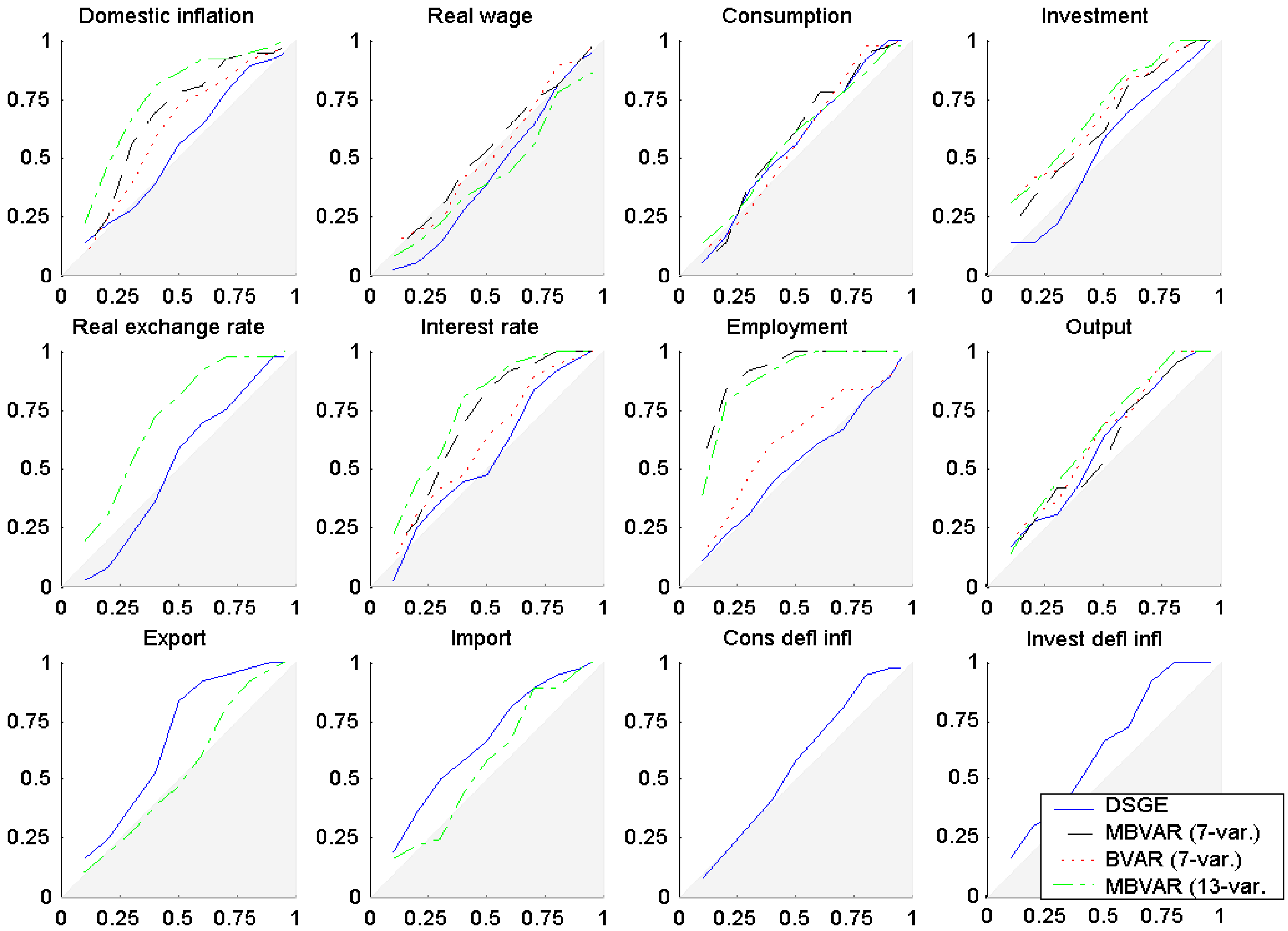


Figure 3b: Empirical coverage probability, DSGE and BVARs 4 quarter horizon

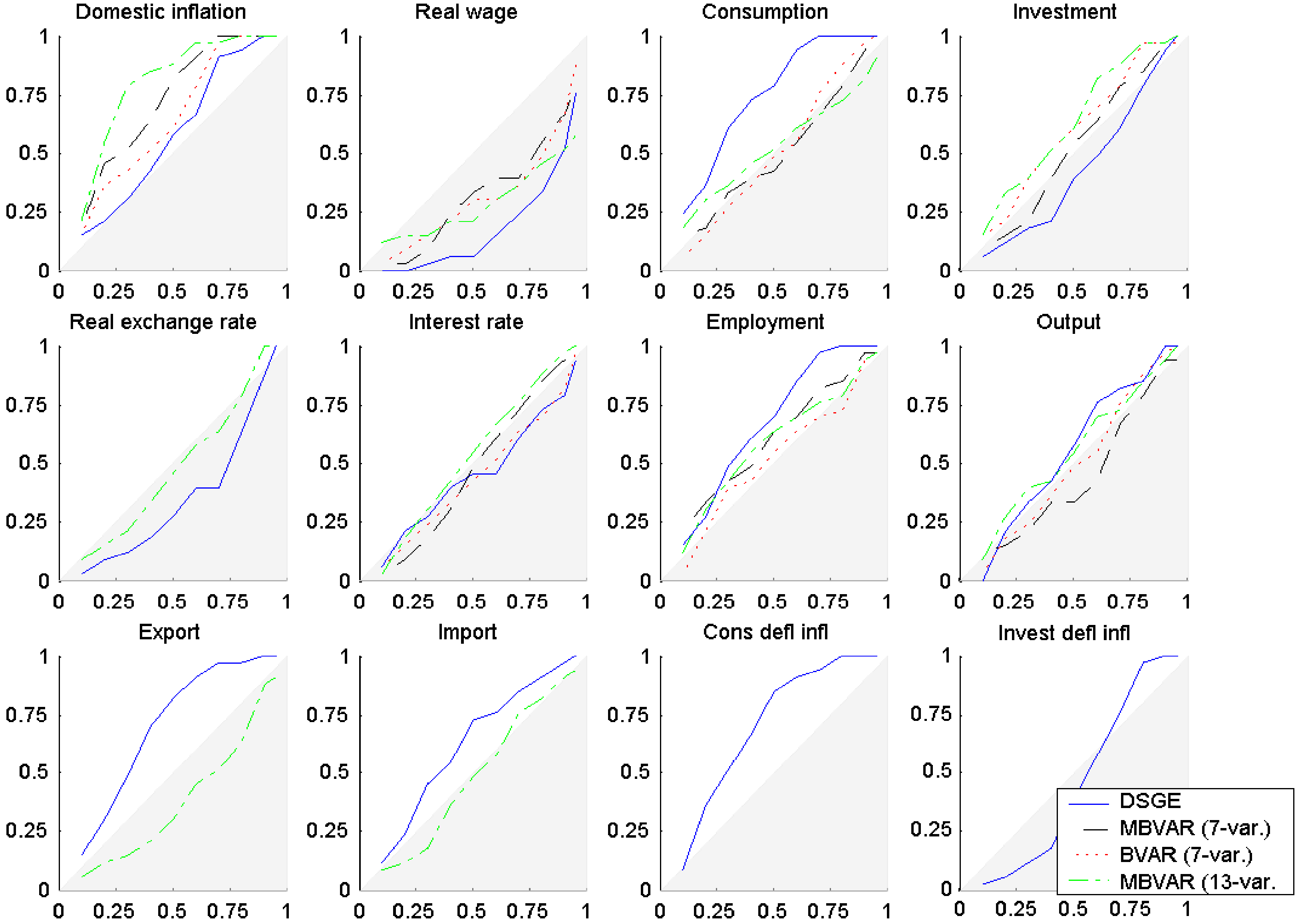


Figure 3c: Empirical coverage probability, DSGE models 1 quarter horizon

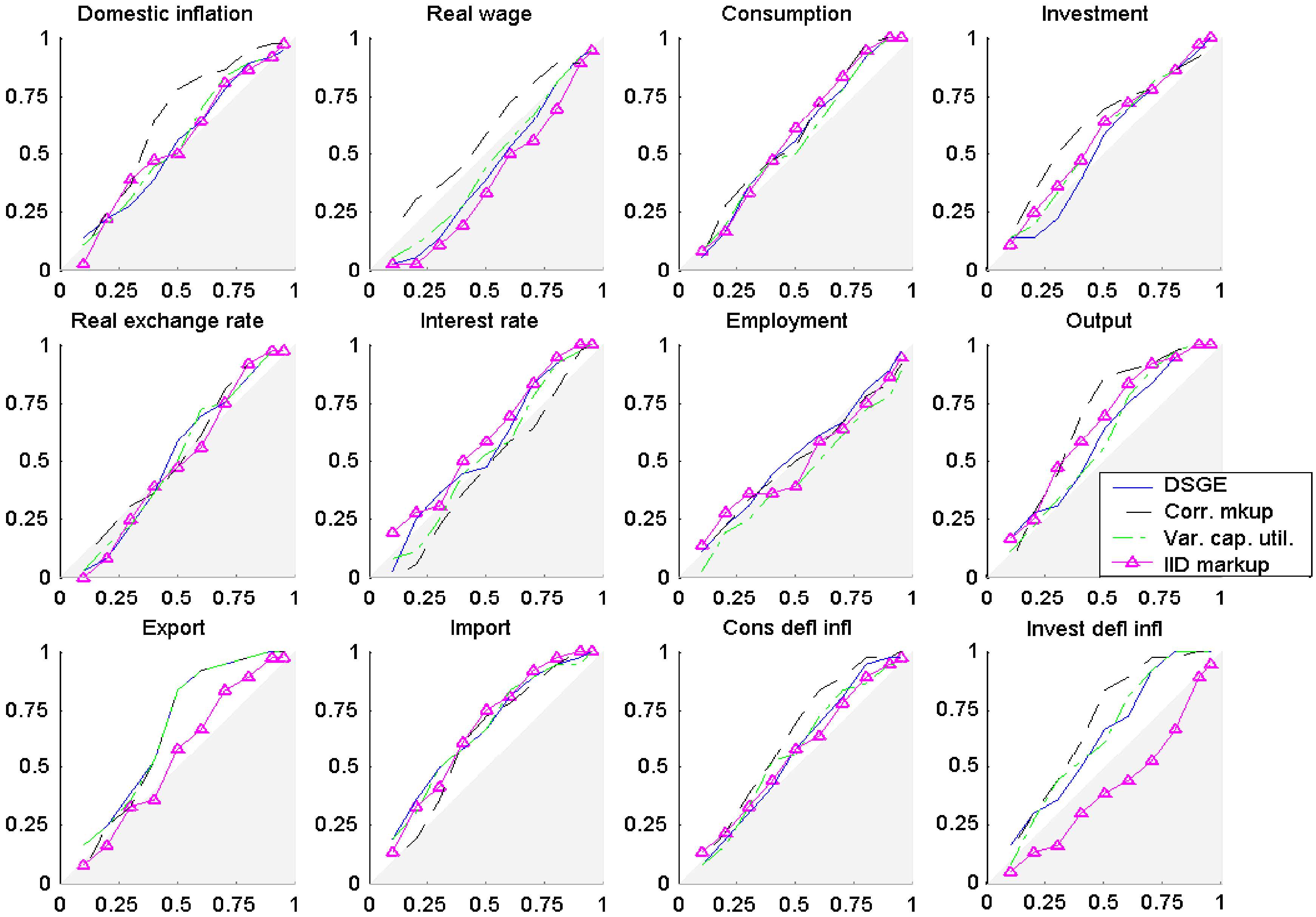
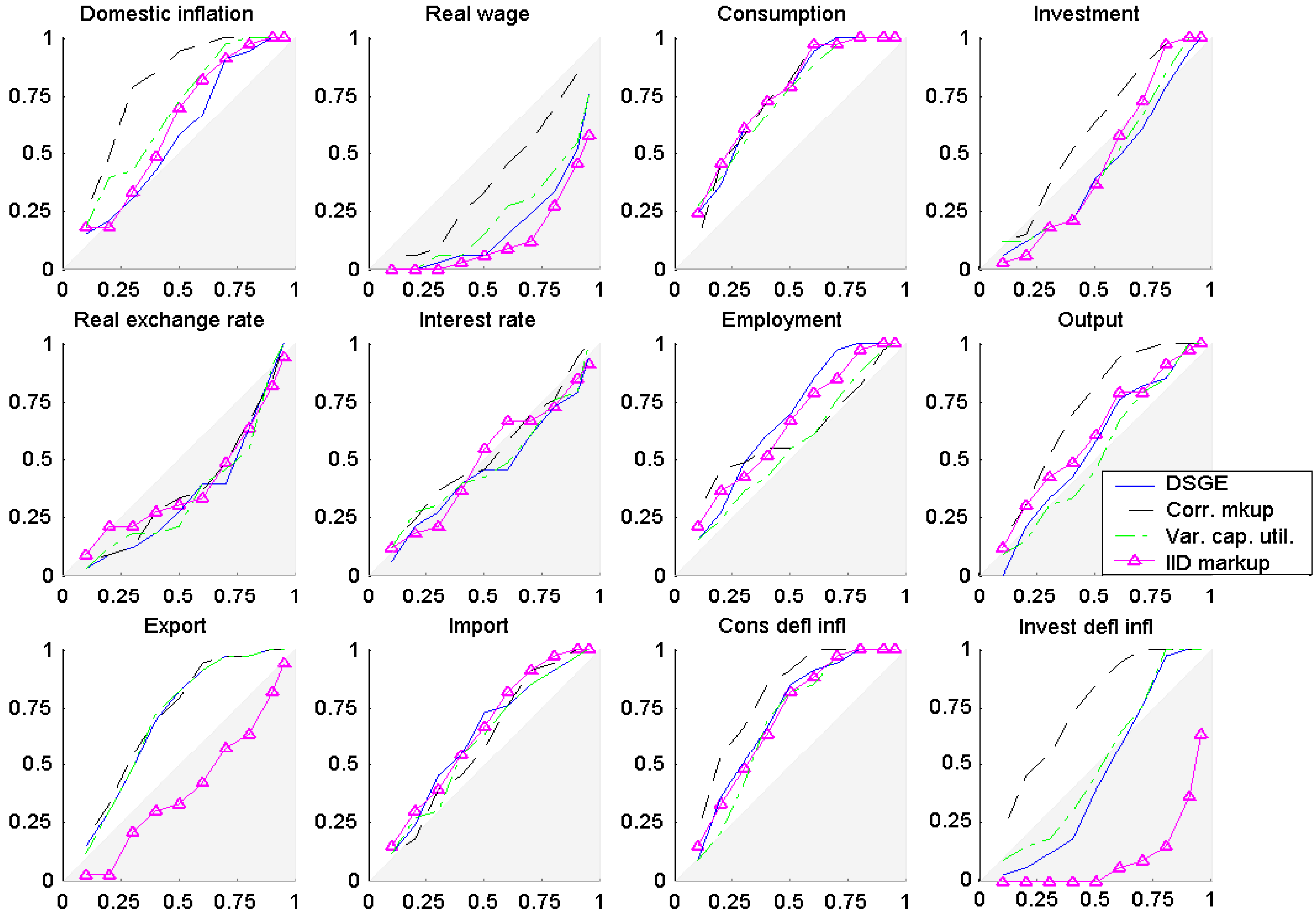


Figure 3d: Empirical coverage probability, DSGE models 4 quarter horizon



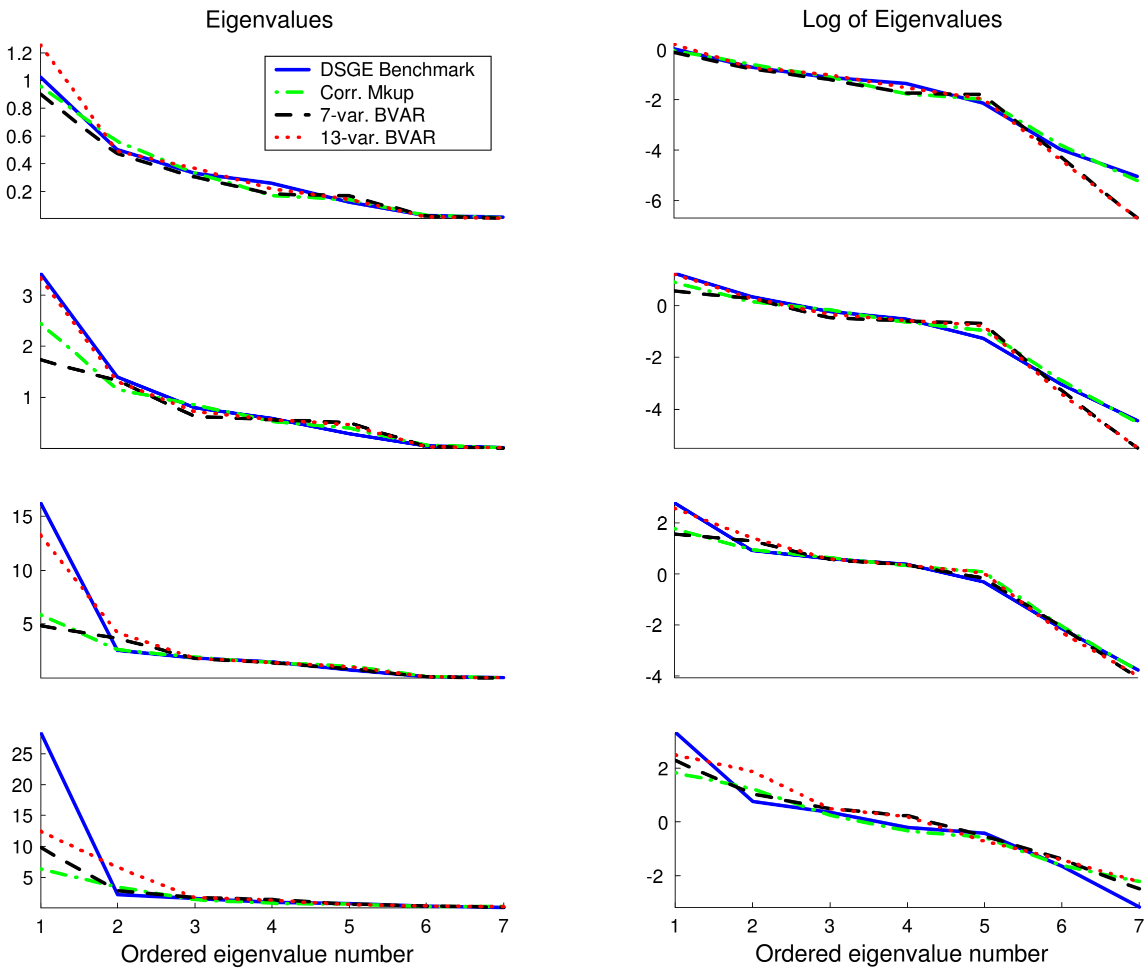


Figure 4: Eigenvalues (left column) and log of eigenvalues (right column) of the MSSE matrix at different horizons.

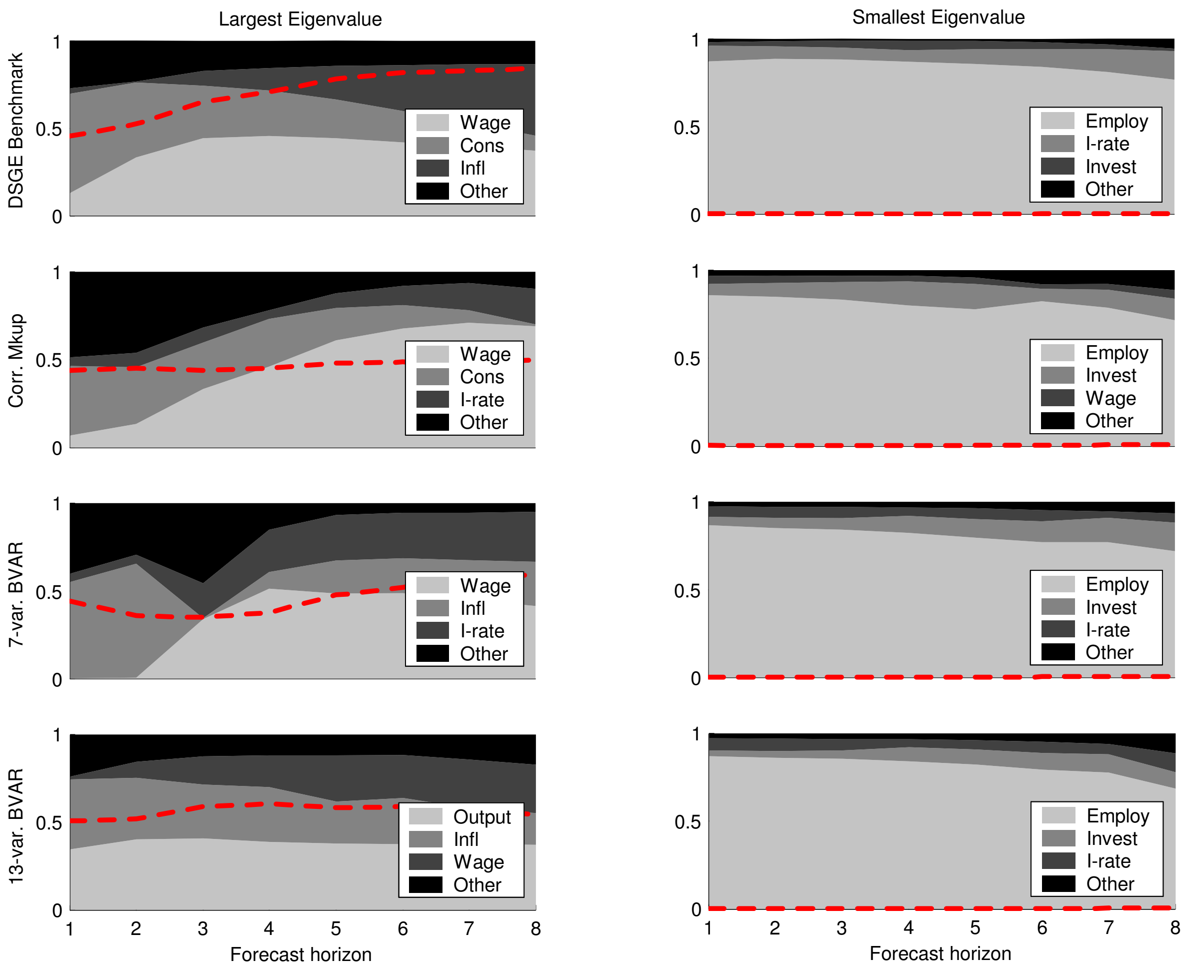


Figure 5: Relative contribution to the largest and smallest eigenvalues of the MSSE matrix (square of the elements in the eigenvectors). The superimposed red dashed line depicts the percentage of total variation explained by the eigenvector, $\lambda_i / \sum_{j=1}^{k} \lambda_j$, where λ_i is the ith largest eigenvalue.