

Putting the New Keynesian DSGE model to the real-time forecasting test*

Marcin Kolasa[†] Michał Rubaszek[‡] Paweł Skrzypczyński[§]

December 27, 2011

Abstract

DSGE models have recently become standard tools for policy-oriented analyses. Nevertheless, their forecasting properties are still barely explored. We address this issue by comparing the quality of real-time forecasts from a richly-specified DSGE model to those from the SPF and DSGE-VARs. We show that the analyzed DSGE model is relatively successful in forecasting the US economy in the period of 1994-2008. Except for short-term forecasts of inflation and interest rates, it is as good as or clearly outperforms DSGE-VARs. Compared to the SPF, the DSGE model generates better output forecasts at longer horizons, but less accurate short-term forecasts for interest rates. Conditional on experts' nowcasts, however, the forecasting power of the DSGE turns out to be similar or better than that of the SPF for all the variables and horizons. An important weakness of the benchmark DSGE model is a poor absolute performance of its point forecasts and rather badly calibrated forecast densities.

Keywords: Forecasting; DSGE; DSGE-VAR; SPF; Real-time data.

JEL Classification: C11; C32; C53; D58; E17

*The earlier versions of this paper were presented at the conference "Publishing Central Bank Forecasts in Theory and Practice" held by the National Bank of Poland (Warsaw 2009); the 64th European Meeting of the Econometric Society (Barcelona 2009); Macroeconomic Forecasting Conference held by IFO, INSEE and ISAE (Rome 2009); the Economic Seminar in the Warsaw School of Economics (Warsaw 2009) and the 42nd Annual Conference of the Money Macro and Finance Research Group (Limassol 2010). We are grateful to participants at these meetings, Michał Brzoza-Brzezina, Wojciech Charemza, Jesus Crespo-Cuaresma, Marco Del Negro, Andrzej Kocięcki, Frank Schorfheide and Tara Sinclair for helpful comments.

[†]National Bank of Poland and Warsaw School of Economics. E-mail: marcin.kolasa@nbp.pl

[‡]National Bank of Poland and Warsaw School of Economics. E-mail: michal.rubaszek@nbp.pl

[§]National Bank of Poland. E-mail: pawel.skrzypczynski@nbp.pl

1 Introduction

The recently observed rise of transparency among central banks in terms of communicating their view on the future state of the economy to the public (see Geraats, 2009) has increased the importance of accurate forecasts in the monetary policy-making process. At the same time, following advancements in the economic and econometric theory, as well as growing computational power of computers, many central banks have started to use dynamic stochastic general equilibrium (DSGE) models for policy-oriented analyses and forecasting. It is well known, however, that in practice central banks usually do not publish pure forecasts generated from their models, but rather make adjustments based on experts' judgment about the future state of the economy. This raises the following two questions, which are the focus of this article. First, how do forecasts generated by DSGE models compare to judgment-based forecasts? Second, how can model-based forecasts be combined with experts' knowledge? For this purpose, we evaluate the relative accuracy of real-time forecasts formulated on the basis of estimated models and by experts. The main question we pose is whether a richly-specified New Keynesian DSGE is able to forecast the US economy better than the Survey of Professional Forecasters (SPF), which we consider to represent the best available judgment-based private sector forecasts. Moreover, we extend our forecasting contest for a relatively new tool for policy-oriented analyses, vector autoregressions using priors from a DSGE model (DSGE-VARs).

The discussion on forecasting properties of DSGE models can generally be divided into two parts. The first strand of the literature uses latest-available data to compare the accuracy of forecasts from DSGE models to that from vector autoregressions (VARs) or their Bayesian versions (BVARs). Smets and Wouters (2007), on the basis of quarterly data for the period 1990:1-2004:4, show that a richly-specified DSGE model is able to outperform VAR and BVAR models in forecasting key macroeconomic variables of the US economy, especially if longer horizons are considered. Del Negro et al.

(2007) develop the DSGE-VAR version of the Smets and Wouters (2003) model (with cointegrating restrictions) and demonstrate that it is able to forecast the US economy better than unrestricted VARs over the evaluation sample of 1985:4-2000:1. In another path-breaking article, Adolfson, Lindé, and Villani (2007) investigate the performance of an open-economy version of the Smets and Wouters (2003) model in forecasting the euro area economy. Using data for the period 1994:1-2002:4 they observe that the accuracy of forecasts from their DSGE model is comparable or even superior to those from VARs and BVARs, both if point forecasts and the whole forecast distributions are considered. All the above articles indicate that the forecasting performance of DSGE models can be better than that of VARs. Consequently, the authors claim that the use of DSGE models in forecasting should increase. We argue that for this statement to be persuasive, DSGE models should also perform well in comparison to judgment-based forecasts.

The second strand of the literature addresses this issue by comparing forecasts from DSGE models to those formulated by experts. It should be noted that in this kind of analysis it is necessary to use real-time data to ensure that information available to experts and estimated models is comparable. To the best of our knowledge, there are only three studies comparing the forecasting performance of DSGE models with judgment-based forecasts in a real-time context. Rubaszek and Skrzypczynski (2008) demonstrate that for the period 1994:1-2006:2 a small-scale DSGE model is able to better forecast GDP growth in the US than the SPF, while it performs relatively poorly in explaining the future paths of inflation and interest rates. Edge, Kiley, and Laforte (2010) compare forecasts from a large-scale DSGE model to those of the Federal Reserve staff and find that in the evaluation sample of 1996:3-2002:4 the forecast accuracy of the DSGE model is superior for real sector variables, and inferior for inflation and interest rates. Finally, Lees, Matheson, and Smith (2007) analyze the accuracy of forecasts formulated by the Reserve Bank of New Zealand (RBNZ) staff relative to those from a

small-scale open economy DSGE model and its DSGE-VAR version. On the basis of the evaluation sample of 1998:4-2003:3 the authors find that the DSGE model is relatively successful in forecasting GDP growth, whereas the RBNZ is doing better in forecasting inflation and interest rates. It should be noted that the precision of forecasts from the DSGE and DSGE-VAR models was found to be comparable. The general picture that emerges from these three articles is that DSGE models perform relatively well in forecasting real sector variables, whereas forecasts for nominal variables are less precise than those formulated by experts.

More recent papers offer additional insights into the predictive power of DSGE models. Edge and Gurkaynak (2010) argue that even though DSGE models can do better than statistical and judgmental forecasts, their absolute forecasting performance is rather poor, especially if evaluated over the period of the Great Moderation, when business cycles were driven mainly by temporary and unforecastable shocks. Herbst and Schorfheide (2011) look into the ability of DSGE models to predict comovements between main macroeconomic variables, documenting some successes but also questioning the usefulness of incorporating into the models additional features improving the empirical fit. Schorfheide, Sill, and Kryshko (2010) show how a DSGE model can be used to generate forecasts of not explicitly modelled variables and find that such forecasts are competitive with simple autoregressive benchmarks.

In this article we add to the literature by investigating the real-time forecasting properties of the Smets and Wouters (2007) DSGE model, which can be considered to represent a benchmark specification for most DSGE models that are currently used in central banks. Our main contribution is fivefold. First, we show that this DSGE model outperforms DSGE-VAR models in forecasting key US macroeconomic variables. Second, we confirm the finding from the literature that, compared to judgment-based forecasts, DSGE models are relatively good in forecasting GDP growth and relatively bad in forecasting interest rates. Third, we indicate that this feature is due to information

advantage of experts: forecasts of nominal variables from the DSGE model, conditional on nowcasts from the SPF, are comparable to forecasts from the SPF. Fourth, we point out that despite a good performance relative to the alternatives, forecasts generated from the DSGE model are still far from satisfactory in the absolute sense. Fifth, we find that the forecast densities implied by the DSGE model are rather poorly calibrated.

The rest of the article is structured as follows. The next three sections present the methods applied to generate the forecasts: the DSGE and DSGE-VAR models, and the SPF. In section 5 we describe the real-time data used in our analysis. Section 6 focuses on parameter estimates and properties of the DSGE and DSGE-VAR models. Section 7 presents the results of the out-of-sample forecast performance analysis. The last section offers conclusions based on the study’s main findings.

2 The DSGE model

The DSGE model proposed by Christiano, Eichenbaum, and Evans (2005) and estimated by Smets and Wouters (2003) using Bayesian techniques is currently considered to be the benchmark richly-specified DSGE model for a closed economy. In this paper we use the Smets and Wouters (2007) version of this model, modified by removing the wage mark-up shock and the wage measurement equation.¹ As the model is well documented in the above-referenced articles, here we only summarize its main structure.

2.1 Final good producers

The final good Y_t is a composite made of a continuum of intermediate goods $Y_{i,t}$ given implicitly as in Kimball (1995) by:

$$1 = \int_0^1 \Gamma \left(\frac{Y_{i,t}}{Y_t}; \lambda_p, \varepsilon_t^p, \varepsilon_p \right) di, \quad (1)$$

¹The reason for this modification is the lack of real-time series for wages in our database.

where Γ is a strictly concave and increasing function such that $\Gamma(1) = 1$. In the formula above, ε_p controls the curvature of the demand price elasticity,² λ_p is the steady-state price mark-up, while ε_t^p is the disturbance to the price mark-up following $\ln \varepsilon_t^p = \rho_p \ln \varepsilon_{t-1}^p + \eta_t^p - \theta_p \eta_{t-1}^p$, $\eta_t^p \sim NID(0, \sigma_p^2)$. The final good producers minimize the cost of producing Y_t , sold at price P_t , by choosing $Y_{i,t}$, each priced at $P_{i,t}$, subject to (1).

2.2 Intermediate goods producers

Intermediate good i is produced using capital services $K_{i,t}^s$ and labor $L_{i,t}$ as inputs according to the technology:

$$Y_{i,t} = \varepsilon_t^a (K_{i,t}^s)^\alpha (\gamma^t L_{i,t})^{1-\alpha} - \Phi_t, \quad (2)$$

where ε_t^a is the productivity disturbance that follows $\ln \varepsilon_t^a = \rho_a \ln \varepsilon_{t-1}^a + \eta_t^a$, $\eta_t^a \sim NID(0, \sigma_a^2)$, and γ represents the deterministic rate of labor-augmenting technological progress. Fixed costs in production Φ_t are related to the steady-state price mark-up through the zero-profit condition $\Phi_t = (\lambda_p - 1)\bar{Y}_t$, where \bar{Y}_t denotes output on the balanced growth path.

In each period a fraction $1 - \xi_p$ of randomly selected firms are allowed to re-optimize. The remaining firms update mechanically according to $P_{i,t} = P_{i,t-1} (\pi_{t-1})^{\iota_p} (\bar{\pi})^{1-\iota_p}$, where $\pi_t = P_t/P_{t-1}$ and $\bar{\pi}$ are the actual and steady-state gross inflation rates.

2.3 Households

Households, indexed by j , choose consumption $C_{j,t}$, hours worked $L_{j,t}$, nominal one-period bond holdings $B_{j,t}$, investment $I_{j,t}$ and capital utilization $Z_{j,t}$ to maximize:

²See Eichenbaum and Fisher (2007) for details.

$$E_t \sum_{s=0}^{\infty} \beta^s \left[\frac{(C_{j,t+s} - hC_{t+s-1})^{1-\sigma_c}}{1-\sigma_c} \right] \exp \left(\frac{\sigma_c - 1}{1 + \sigma_l} L_{j,t+s}^{1+\sigma_l} \right), \quad (3)$$

subject to the nominal budget constraint:

$$\frac{B_{j,t}}{R_t \varepsilon_t^b} \leq B_{j,t-1} + W_{j,t} L_{j,t} + R_t^k Z_{j,t} K_{j,t-1} + Div_t - P_t (T_t + C_{j,t} + I_{j,t} + \Psi(Z_{j,t}) K_{j,t-1}) \quad (4)$$

and the capital accumulation equation:

$$K_{j,t} = (1 - \delta) K_{j,t-1} + \varepsilon_t^i \left[1 - S \left(\frac{I_{j,t}}{I_{j,t-1}} \right) \right] I_{j,t}. \quad (5)$$

Here, Div_t denotes dividends received from firms, T_t stands for lump-sum net taxes, W_t is the nominal wage and R_t^k is the rental rate on capital. The rate of return on assets held by households is a product of the gross interest rate set by the central bank R_t and the risk premium disturbance ε_t^b that follows $\ln \varepsilon_t^b = \rho_b \ln \varepsilon_{t-1}^b + \eta_t^b$, $\eta_t^b \sim NID(0, \sigma_b^2)$. The parameters h , σ_c , σ_l and δ represent external habit formation, the inverse of the intertemporal elasticity of substitution, the inverse of the Frisch elasticity and the capital depreciation rate, respectively.

The accumulation of capital $K_{j,t}$ is subject to adjustment costs given by a function S that satisfies $S(\gamma) = 0$, $S'(\gamma) = 0$ and $S''(\gamma) = \varphi$. It also depends on the investment-specific productivity disturbance following $\ln \varepsilon_t^i = \rho_i \ln \varepsilon_{t-1}^i + \eta_t^i$, $\eta_t^i \sim NID(0, \sigma_i^2)$. The accumulated capital is subsequently transformed into capital services $K_{j,t}^s = Z_{j,t} K_{j,t-1}$ that are sold to firms. Finally, households have to pay real costs of capital utilization $\Psi(Z_{j,t}) K_{j,t-1}$, where Ψ is an increasing function that satisfies $\Psi(1) = 0$ and $\frac{\Psi''(1)}{\Psi'(1)} = \frac{\psi}{1-\psi}$.

2.4 Labor market

Labor supplied by individual households $L_{j,t}$ is combined into aggregate labor L_t by perfectly competitive labor packers according to the Kimball (1995) formula:

$$1 = \int_0^1 \Gamma \left(\frac{L_{j,t}}{L_t}; \lambda_w, \varepsilon_w \right) dj, \quad (6)$$

where λ_w represents the steady-state wage mark-up and ε_w characterizes the curvature of the labor demand elasticity. The aggregated labor is subsequently sold to intermediate goods producers at price W_t . The labor packers minimize the cost of generating L_t , subject to (6).

Wage setting is subject to nominal rigidities á la Calvo, which means that in each period only a fraction $1 - \xi_w$ of households are allowed to re-optimize their wages. The remaining households adjust their wages mechanically according to $W_{i,t} = W_{i,t-1} \gamma (\pi_{t-1})^{\iota_w} (\bar{\pi})^{1-\iota_w}$.

2.5 Closing the model

The central bank follows a generalized Taylor rule:

$$\frac{R_t}{\bar{R}} = \left(\frac{R_t}{\bar{R}} \right)^\rho \left[\left(\frac{\pi_t}{\bar{\pi}} \right)^{r_\pi} \left(\frac{Y_t}{Y^p} \right)^{r_y} \right]^{1-\rho} \left(\frac{Y_t/Y_{t-1}}{Y_t^p/Y_{t-1}^p} \right)^{r_{\Delta y}} \varepsilon_t^r, \quad (7)$$

where \bar{R} denotes the steady-state nominal interest rate, Y_t^p is potential output defined as in Smets and Wouters and ε_t^r is the monetary policy shock that follows $\ln \varepsilon_t^r = \rho_r \ln \varepsilon_{t-1}^r + \eta_t^r$, $\eta_t^r \sim NID(0, \sigma_r^2)$.

Government spending G_t is driven by an exogenous process $G_t = g_y Y_t \varepsilon_t^g$, where g_y denotes the steady-state share of government purchases in output and $\ln \varepsilon_t^g = \rho_g \ln \varepsilon_{t-1}^g + \eta_t^g + \rho_{ga} \eta_t^a$, $\eta_t^g \sim NID(0, \sigma_g^2)$.

The model is closed by the aggregate resource constraint of the following form:

$$Y_t = C_t + I_t + G_t + \Psi(Z_t)K_{t-1}. \quad (8)$$

2.6 Solution and estimation

The empirical implementation of the DSGE model can be described as follows. First, the model is linearized around its steady-state and written as a linear expectation system. The linearized version is described in detail by Smets and Wouters (2007), with the difference that we do not include the wage mark-up disturbance. For given parameter values, such a system can be solved out using standard techniques and transformed into a state-space representation, where the measurement equations relate the model variables to macroeconomic data. Having this representation, the likelihood of the model can be obtained with the Kalman filter.

The structural parameters of the DSGE model are estimated by applying Bayesian techniques. Our assumptions for the priors and five calibrated parameters are identical to those used by Smets and Wouters. For each sample, the posterior mode and the corresponding Hessian matrix are calculated using standard numerical optimization routines. The posterior distribution is approximated using the Metropolis-Hastings algorithm with 125,000 replications, out of which we drop the first 25,000.

3 DSGE-VAR models

It is well known that a standard DSGE model has a restricted infinite-order VAR representation.³ Therefore, VARs have been widely used in the literature as unconstrained benchmarks for evaluating DSGE models. However, because of the large number of parameters and short time series, estimates of unrestricted VAR coefficients are in many cases imprecise and forecasts have large standard errors. As it is common in the literature, we tackle this problem by using a Bayesian approach. Following Del Negro and Schorfheide (2004), we consider a theoretical prior that is based on the DSGE model described in the previous section.

³See Fernández-Villaverde et al. (2007) for sufficient conditions regarding the VAR representation of a DSGE model.

More specifically, we analyze a VAR model:

$$z_t = A_0 + \sum_{i=1}^p A_i z_{t-i} + u_t, \quad (9)$$

where z_t is an n -dimensional vector of observed variables, A_i are matrices of model coefficients, $u_t \sim NID(0, \Sigma_u)$ is the error term, and p denotes the maximum lag order.

Model (9) can be expressed in the matrix form as:

$$Z = XA + U, \quad (10)$$

where Z is the $T \times n$ matrix with rows z'_t , X is the $T \times (np + 1)$ matrix with rows $x_t = [1, z'_{t-1}, \dots, z'_{t-p}]$, U is the $T \times n$ matrix with rows u'_t , $A = [A_0, A_1, \dots, A_p]'$ and T is the sample size. The likelihood function, conditional on observations x_{1-p}, \dots, x_0 , can be expressed as:

$$f(Z|A, \Sigma_u) \propto |\Sigma_u|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr} [\Sigma_u^{-1} (Z - XA)' (Z - XA)] \right\}. \quad (11)$$

A DSGE-VAR model can be thought of as a result of adding λT artificial observations simulated from the DSGE model to the actual data and estimating the VAR model on the basis of a mixed sample of the artificial and actual observations. The hyperparameter λ denotes the prior tightness so that for $\lambda = 0$ the DSGE-VAR model corresponds to the unrestricted VAR and for $\lambda = \infty$ the DSGE-VAR model becomes the VAR representation of the DSGE model.

A short description of the Del Negro and Schorfheide procedure is as follows. Given the parameters of the DSGE model θ and its state-space representation, it is possible to compute the expected values of artificial data sample moments: $\Gamma_{zz}^* = E_\theta(z_t z'_t)$, $\Gamma_{zx}^* = E_\theta(z_t x'_t)$, $\Gamma_{xz}^* = E_\theta(x_t z'_t)$ and $\Gamma_{xx}^* = E_\theta(x_t x'_t)$. The conjugate prior of the VAR coefficients Σ_u and A conditional on θ is of the Inverse Wishart-Normal form:

$$\begin{aligned}
\Sigma_u | \theta, \lambda &\sim IW \left(\lambda T \left(\Gamma_{zz}^* - \Gamma_{zx}^* \Gamma_{xx}^{*-1} \Gamma_{xz}^* \right); \lambda T - (np + 1) \right) \\
A | \Sigma_u, \theta, \lambda &\sim N \left(\Gamma_{xx}^{*-1} \Gamma_{xz}^*; \Sigma_u \otimes (\lambda T \Gamma_{xx}^*)^{-1} \right).
\end{aligned} \tag{12}$$

This means that the posterior distribution of the VAR coefficients is:

$$\begin{aligned}
\Sigma_u | Z, \theta, \lambda &\sim IW \left((\lambda + 1) T \hat{\Sigma}_u; (\lambda + 1) T - (np + 1) \right) \\
A | Z, \Sigma_u, \theta, \lambda &\sim N \left(\hat{A}; \Sigma_u \otimes (\lambda T \Gamma_{xx}^* + X'X)^{-1} \right),
\end{aligned} \tag{13}$$

where $\hat{A} = (\lambda T \Gamma_{xx}^* + X'X)^{-1} (\lambda T \Gamma_{xz}^* + X'Z)$ and $\hat{\Sigma}_u = [(\lambda + 1) T]^{-1} [\lambda T (\Gamma_{zz}^* - \Gamma_{zx}^* \hat{A} + (Z'Z - Z'X \hat{A}))]$. One can notice that the expected value of the posterior distribution of the VAR parameters A is a weighted average of the estimates implied by the expected moments from the DSGE model and the unrestricted OLS estimates, where the weight is determined by the hyperparameter λ .

As in Del Negro and Schorfheide, the prior assumptions given by (12) are complemented with a prior distribution of DSGE model parameters. Following Adjemian, Paries, and Moyen (2008), we also define a prior distribution for the hyperparameter λ , which is assumed to be uniform over the interval $[0,10]$. The VAR coefficients and the parameters related to the DSGE model, including λ , are estimated jointly as the posterior distribution is factorized into the posterior density of the former given the latter and the marginal posterior density of the latter. As in the case of the DSGE model, the posterior distribution of DSGE-VAR parameters is obtained numerically using standard optimization routines and the Metropolis-Hastings algorithm with 125,000 replications, out of which we drop the first 25,000.

4 The Survey of Professional Forecasters

The SPF is the oldest quarterly survey of macroeconomic forecasts in the United States. The survey, which was launched and elaborated by the American Statistical Association and the National Bureau of Economic Research in 1968, was taken over by the Federal Reserve Bank of Philadelphia in 1990.⁴ It is carried out at regular three-month intervals and concerns dozens of macroeconomic variables, among them real GDP, the GDP price index and the three-month Treasury bill (TB) rate. In the further part of this paper we focus on the median forecasts of the above-listed variables by the SPF, which concern the period when the survey is carried out (so they are essentially nowcasts) and the next four quarters.

As discussed in more detail by Croushore (2010), the survey's forms are sent at the end of the first month of each quarter, just after the advance release of the national account data for the previous period. The respondents return them in the middle of the next month, i.e. before the data are revised. Nevertheless, the forecasters may use some additional information while formulating their predictions for the US economy, in particular if they monitor leading indicators, business surveys or developments in financial markets. Therefore, it seems obvious that the SPF has an advantage in forecasting, and even more in nowcasting output, prices and especially interest rates in comparison to the estimated models described above. We will address this issue in the second part of our forecasting accuracy investigation.⁵ On the other hand, as pointed out by Edge, Kiley, and Laforte (2010), the DSGE and DSGE-VAR models have an advantage over the SPF in retrospective forecasting of the US economy as these models benefit from the research on what types of models are well fitted to the data. For example, neither the structure of the Smets and Wouters model nor the priors used in Bayesian

⁴The results of the survey are published quarterly on the Philadelphia Fed website: <http://www.phil.frb.org/econ/spf/index.html>.

⁵Yet another advantage of the SPF, implied by the forecast averaging literature, is that the median forecaster is not the same for each forecasting round, variable and horizon.

estimation were available two decades ago, i.e. in time of forecast formulation by the SPF. Unfortunately, it seems impossible to control our results for this kind of potential biases.

5 The data

The DSGE and DSGE-VAR models were estimated on the basis of six key US macroeconomic variables: real GDP, real consumption, real investment, the GDP price index (all expressed in log-differences), log hours worked, and the three-month TB rate. Since the use of the latest-available data in the estimation would give an advantage to the estimated models over the SPF in ex-post forecasts comparisons due to data revisions, we applied the Philadelphia Fed “Real-Time Data Set for Macroeconomists”, which is described in more detail by Croushore and Stark (2001). This ensures the comparability of the forecasting errors as all predictions are formulated using a similar data set.

The out-of-sample forecast performance is analyzed for horizons ranging from zero up to four quarters ahead, whereas the evaluation is based on the data from the period 1994:1-2008:4, called henceforth the evaluation sample. The DSGE and DSGE-VAR models were estimated on the set of the recursive samples starting in 1964:2 and ending one quarter before a given vintage date, which is the period of forecast formulation. For instance, the forecasts elaborated in 1994:1 for the period 1994:1-1995:1 were generated using the models estimated on the basis of observations from 1964:2 to 1993:4, using the data available in 1994:1. This procedure is repeated for each quarter from the period 1994:1-2007:4, which gives 56 forecasts for each forecast horizon, model and variable.

6 Recursive estimates of model parameters

Before we move to evaluation of forecasts generated by our competing models, we first briefly discuss the estimation results obtained for the DSGE model. The distribution

characteristics of the 1964:2-1993:4 to 1964:2-2008:4 recursive estimates of the posterior median are reported in the right-side columns of Table 1. For convenience, in the left-side columns we also report our assumptions for the priors and five calibrated parameters. They are identical to those used by Smets and Wouters. The only additional parameter we use is the trend population growth rate, which we calibrate at 1.5% per annum and include in the measurement equations for output, consumption and investment.⁶

Comparing our recursive estimates with those obtained by Smets and Wouters we note the following major differences. Averaged across all samples, our results point at lower costs of adjusting investment φ and capacity utilization ψ , less persistent habits h , higher labour supply elasticity σ_l , a higher steady-state price mark-up λ_p and a lower trend growth rate γ . There are also some differences in the characteristics of the shock processes.⁷ The average medians for the remaining parameters fall within the 90% confidence intervals reported by Smets and Wouters. A closer inspection reveals that these discrepancies are almost entirely due to data and sample differences. As we find out by experimenting with the original dataset used by Smets and Wouters, dropping the wage mark-up shock from the model and real wages from the set of observable variables leads only to significantly higher estimates of the consumption elasticity σ_c and the price mark-up shock inertia ρ_p .⁸

We proceed by evaluating the model's stability and the speed of reversion to the steady-state by looking at the recursive median impulse response functions (IRF) to the structural shocks. An informal analysis of Figure 1 shows that the model is relatively

⁶The reason for this correction is that the real-time series we use are not expressed in per capita terms as in the Smets and Wouters paper. Our results are robust to alternative (reasonable) calibrations of this parameter.

⁷The results reported in Table 1 point at an apparently high estimate of the risk premium volatility. However, as noted by Taylor and Wieland (2009), the risk premium volatility estimated by Smets and Wouters (2007) is actually multiplied by the interest elasticity of consumption.

⁸Similarly to Guerron-Quintana (2010), we find that omitting real wages lowers the estimate of wage stickiness ξ_w and increases the estimate of price stickiness ξ_p . These changes, however, are not sizable.

stable over the evaluation sample. Importantly, despite the above mentioned differences in the estimates of some of the parameters, the impulse responses turn out to be very similar to those reported in Smets and Wouters.

The recursive estimates of the DSGE-VAR hyperparameter λ are presented in Table 2. According to the results for the DSGE-VAR model with two lags, the average median recursive estimate of λ is relatively low, indicating that the data give 36.4% probability to the VAR representation of the DSGE model and 63.6% probability to the unrestricted VAR. This result should not be surprising as the persistence embedded in our benchmark DSGE model can be approximated by a VAR process with two lags only to a limited degree. Increasing the maximum lag p of the DSGE-VAR model leads to a rise in the estimated tightness of the DSGE-based prior. In particular, for $p = 8$ the average posterior estimate of λ indicates 66.4% probability of the DSGE model.

In the next section, we apply the DSGE and DSGE-VAR models to forecasting the US economy. For this purpose, we generate out-of-sample density forecasts that take into account uncertainty related to model parameters and stochastic shocks. We repeat this procedure for each quarter from the evaluation sample. All calculations are performed with the DYNARE package for MATLAB 7.

7 Forecasts comparison

Good forecast accuracy is one of the key criteria in the process of model evaluation. In this section we compare the forecasting performance of the DSGE model to that of the SPF and several variants of DSGE-VARs, differing in the number of lags in the VAR approximation.⁹ We use two main statistics commonly applied for evaluation of point forecasts: the mean forecast error (MFE) and the root mean squared forecast error (RMSFE). We complement them with efficiency tests that give us an idea about each

⁹In the working paper version of this article we also reported the results for atheoretical BVARs with Minnesota priors. They are broadly consistent with those obtained from the DSGE-VARs.

method’s absolute forecasting performance. Finally, we evaluate density forecasts. All calculations are done using the “actuals” taken from the last vintage of our sample, but we found that the results with different “actuals” are broadly the same.

7.1 Mean forecast errors

We begin our forecasting contest by investigating the MFEs for three key US macroeconomic variables: output growth, inflation and the interest rate. The forecast horizon h ranges from zero (nowcasts) to four quarters (maximum horizon of the SPF forecasts).

According to the results presented in Table 3, output growth forecasts are unbiased only for the DSGE model and the SPF. The DSGE-VAR models tend to significantly overpredict the future path of GDP for all forecast horizons and for all values of the maximum lag p . As regards inflation forecasts, all methods perform quite well: the MFEs are not significantly different from zero. Finally, most methods overestimate the future level of the interest rate and the bias is increasing with the forecast horizon h . Overall, the results indicate that, as far as the MFEs are concerned, the SPF and the DSGE dominate the DSGE-VAR models in forecasting output, whereas all methods perform comparably well in the case of inflation and the interest rate.

7.2 Root mean squared forecast errors

We continue our contest by comparing the second moments of the forecast errors. Given the main focus of this paper, in Table 4 we report the levels of the RMSFEs only for the DSGE model. The remaining numbers are expressed as the ratios so that the values above unity indicate that the DSGE model dominates the alternative method. Moreover, we test whether this difference in the RMSFEs is statistically significant using the HLN-DM test (Harvey, Leybourne, and Newbold, 1997), where the long-run variance is calculated with the Newey and West (1987, 1994) method.

According to our results, the accuracy of output growth forecasts from the DSGE

model is significantly higher than that from the remaining methods at most horizons. In comparison to the SPF, the RMSFEs from the DSGE model are about 15 percent lower for two-, three- and four-quarter ahead forecasts. Moreover, the precision of output growth forecasts from the DSGE model is about 20 percent higher than that obtained from the DSGE-VAR models. As far as inflation forecasts are concerned, the RMSFEs from the DSGE model, the SPF, and the DSGE-VAR models with at least 4 lags are comparable, while the lowest order DSGE-VAR performs significantly worse. Finally, interest rate forecasts formulated by the SPF are substantially better than those generated by the estimated models. The SPF dominance is most evident for nowcasts and one-quarter-ahead forecasts.

What is probably most striking in our results is the finding that the RMSFEs from the DSGE model are at least as low as those from the DSGE-VAR models for most variables and horizons. This may be viewed as contrasting with Del Negro et al. (2007), who conclude that DSGE-VAR models with optimally chosen weights of DSGE priors (λ) perform better than their counterparts with $\lambda = \infty$, i.e. with dogmatic DSGE restrictions. However, as demonstrated by Chari, Kehoe, and McGrattan (2008), a VAR with a small number of lags is usually a poor approximation to a DSGE model's infinite-order VAR representation. As can be seen from the last column of Table 4, even if we use the number of lags as large as sixteen, i.e. far more than in any macroeconomic applications, the DSGE-VAR with $\lambda = \infty$ performs worse than the state-space representation of the underlying DSGE model. In other words, gains from relaxing the dogmatic DSGE restrictions turn out to be more than offset by losses related to the lag truncation bias.

Our results also confirm the finding from the earlier literature that, compared to judgment-based forecasts, DSGE models are relatively good in forecasting GDP growth, especially at longer horizons, and relatively weak in forecasting interest rates, especially at shorter horizons. In section 7.4 we show that the relative success of the SPF in short-

term forecasting might be due to an information advantage.

7.3 Unbiasedness tests

The relatively good performance of DSGE forecasts does not imply that they are satisfactory in the absolute sense. To shed some light on this issue, we run a standard forecast unbiasedness test (see Clements and Hendry, 1998). For each forecasting method, variable and forecast horizon, we regress the actuals (x_τ) on forecasts (x_τ^F):

$$x_\tau = \alpha_0 + \alpha_1 x_\tau^F + \eta_\tau. \quad (14)$$

For a good forecast, the constant term should be zero, the slope coefficient equal to one and the model fit should be high. Table 5 reports the estimates of (14) for the DSGE, SPF and the best performing DSGE-VAR, which is, given the MFEs and RMSFEs discussed above, the model with four lags. To save space, we restrict our attention to $h = \{0, 2, 4\}$. For testing forecast unbiasedness, we use the Wald test for the null of $\alpha_0 = 0$ and $\alpha_1 = 1$. All statistics are adjusted for heteroskedasticity and autocorrelation of the residuals with the Newey and West (1987, 1994) procedure.

Only DSGE-based output growth forecasts turn out to be unbiased, while only the SPF produces unbiased nowcasts of this variable. As regards inflation, all methods generate biased forecasts. Finally, the DSGE-VAR performs best at predicting the interest rate, while the DSGE does worst in this area. However, if we look at the R^2 coefficients, the forecasts of output growth and inflation turn out to be very poor by all methods. This finding is illustrated in Figure 2, which plots the forecasts against the realizations. The points form dispersed clouds rather than 45-degree lines. A somewhat better picture emerges for the interest rate, where the fit to actuals is particularly good for nowcasts. However, this result can be attributed to high persistence of the variable. Overall, our findings support those obtained by Rubaszek and Skrzypczynski (2008) and Edge and Gurkaynak (2010), who conclude that while DSGE models are competitive

relative to alternatives, their absolute forecasting performance is rather poor.

7.4 Conditioning on SPF nowcasts

We have already mentioned that the SPF forecasts, and nowcasts in particular, have an advantage over the estimated models as the surveyed experts can observe monthly data for CPI inflation, industrial production, retail sales or leading indicators, which might help in nowcasting output growth and inflation. However, the biggest advantage is for the interest rate as SPF experts know its path up to the middle of the quarter in which forecasts are formulated, whereas the models are estimated with the data ending in the previous quarter. Consequently, it should come as no surprise that the short-run SPF forecasts have the lowest RMSFEs among all investigated methods, and the superiority of the SPF is most evident in the case of the interest rate. Below we address this issue by comparing the accuracy of forecasts from the estimated models that take into account nowcasts formulated by the SPF.¹⁰

We first condition on the SPF nowcasts for all variables used in estimation.¹¹ The results reported in Table 6 show that the DSGE model still significantly outperforms the other methods in forecasting output growth. In the case of inflation, the DSGE and DSGE-VAR models with four or more lags are characterized by the lowest RMSFEs, the SPF is insignificantly less accurate, while the low-order DSGE-VAR is found to be the worst. Finally, the RMSFEs for interest rate forecasts formulated by all methods now do not differ in a significant way.

As we have noted, the information advantage of the SPF concerns particularly the interest rates. Hence, we also run our forecasting contest using the SPF nowcasts of the

¹⁰More specifically, we do not reestimate the models using the SPF nowcasts as actual observations. Doing so made no perceptible difference.

¹¹Since the SPF does not forecast average and total hours worked, we applied bridge regressions based on the SPF nowcasts for GDP to generate nowcasts for total hours worked and payroll employment (the available SPF nowcasts for the latter start in late 2003). Subsequently, we calculated nowcasts for average hours worked as the ratio of the resulting estimates for total hours worked and payroll employment. The results of these regressions are available upon request.

interest rate only.¹² The results reported in Table 7 show that such a partial conditioning is sufficient to make the interest rate forecasts statistically indistinguishable from each other. It also preserves the superiority of DSGE-based output growth forecasts over the medium horizon. It is worth noting, however, that the RMSEs for output nowcasts obtained from the DSGE are now by nearly 20% larger than in the unconditional variant (see 4). This finding is not surprising given the MFEs reported in Table 3, which show that the DSGE tends to overpredict both output growth and the interest rate, and the fact that the model implies a negative comovement between these two variables.¹³

The results discussed above suggest that the superior performance of experts in forecasting nominal variables found in the earlier literature can be attributed to their information advantage, related to having access to high frequency data. Conditional on the SPF nowcasts, the DSGE model is found to outperform the SPF in forecasting output growth and to generate a similar precision of forecasts for inflation and the interest rate. This indicates that including experts' nowcasts in the process of forecast formulation with an estimated structural model can improve the forecast precision. However, our results also show that point forecasts for the remaining horizons should not necessarily be corrected by experts.

7.5 Density forecasts

We complement the point forecast evaluation discussed above with an assessment of density forecasts. The aim is to check to what extent the analyzed forecasts provide a realistic description of actual uncertainty. To this effect we use predictive densities implied by the DSGE and DSGE-VAR models. Since the quarterly SPF does not include information on the forecast uncertainty, we exclude it from the analysis presented in

¹²The conditioning is done with the interest rate shock η_t^r .

¹³Our results are also consistent with Schorfheide, Sill, and Kryshko (2010), who use a variant of the Smets and Wouters (2003) model and report that conditioning on the interest rate increases the RMSFEs for other variables, including GDP growth.

this subsection.¹⁴

We first evaluate the quality of density forecasts with Probability Integral Transform (PIT), developed by Rosenblatt (1952) and imported to the economic literature by Diebold, Gunther, and Tay (1998). The PIT is defined as the transformation:

$$p_\tau = \int_{-\infty}^{x_\tau} f(u)du, \quad (15)$$

where $f(u)$ is the ex-ante forecast density and x_τ is the ex-post realization. If the density forecast is well calibrated, p_τ should be independently and uniformly distributed on the interval (0,1). Diebold, Gunther, and Tay advocate a variety of graphical approaches to forecast evaluation. For instance, one can divide the unit interval into J subintervals and check if the fraction of PITs in each of them is close to J^{-1} . This way of visualization has been recently used for evaluation of DSGE models by Herbst and Schorfheide (2011).

We follow this line, set $J = 10$ and present in Figure 3 the histograms of PITs based on the DSGE and the DSGE-VAR with four lags. We focus on 4-quarter-ahead forecasts. It is clear that for both models GDP growth density forecasts are too diffuse as too many PITs fall into the middle bins. A similar picture emerges for inflation. As regards the interest rate, an unproportional number of PITs fall in the lowest bin, reflecting the fact that both models tend to overpredict the interest rate level.

Overall, the density forecasts from the DSGE-VAR appear somewhat better calibrated. This is because DSGE models impose tight restrictions on the data and hence its misspecification needs to be absorbed by stochastic shocks (Del Negro and Schorfheide, 2009). Since DSGE-VARs relax these restrictions, the estimated variance of shocks is lower, which results in somewhat tighter predictive distributions.

A formal way of testing whether the density forecast are well calibrated is the Berkowitz (2001) test with the null that p_τ is iid and distributed uniformly on (0,1). Berkowitz claims that it is difficult to test for uniformity with small data samples and

¹⁴Interval forecasts from the SPF are available for annual data only.

therefore proposes to analyze the series $z_\tau = \Phi^{-1}(p_\tau)$, where Φ is the cdf of the standard normal distribution. If the null is true, then z_τ should be iid and normally distributed.

To test it, one needs to estimate the following regression

$$q_\tau = \gamma_0 + \gamma_1 q_{\tau-1} + \eta_\tau, \eta_\tau \sim N(0, \sigma_\eta^2), \quad (16)$$

A good density forecast should have an intercept and a slope coefficient of zero and a variance of the error term equal to one.

The estimates of (16) and the likelihood ratio test statistics for the null of $\gamma_0 = 0$, $\gamma_1 = 0$ and $\sigma_\eta = 1$ are presented in Table 8. The following observations can be made. First, except for inflation nowcasts, the model-based density forecasts are badly calibrated. As could be seen by inspecting the PIT histograms, the DSGE-VAR performs somewhat better than the DSGE, but for both models the LR statistics are higher than the critical value at any conventional significance level. Second, the poor calibration of all forecast densities is mainly due to their excessive dispersion (σ_η is below unity in all cases). For the interest rate forecasts, the PITs are also autocorrelated (γ_1 is larger than zero).

8 Conclusions

In this paper we have shown that the overall accuracy of real-time forecasts generated by a richly-specified DSGE model is comparable to judgment-based forecasts formulated by the SPF. While the SPF seems to have a clear advantage in short-term point forecasts of the interest rate, the DSGE model performs significantly better in forecasting GDP at longer horizons. Moreover, the DSGE model has been found to outperform the DSGE-VAR models in forecasting the US economy. We have also demonstrated that the dominance of experts in forecasting nominal variables found in the earlier literature can be attributed to an information advantage of experts, namely their access to current

high-frequency data. Conditional on experts' nowcasts, the RMSFEs from the DSGE model turned out to be comparable or even smaller than the RMSFEs of the SPF forecasts. It has to be stressed, however, that even though the benchmark DSGE model is found to be competitive with other methods, its absolute performance is far from satisfactory. Also, the forecast densities it generates are rather poorly calibrated.

We believe that the above findings contribute to the current discussion on the usefulness of DSGE models in policy oriented analyses. Del Negro et al. (2007) point at an improved time series fit of DSGE models as an important factor behind their increasing use in policy making institutions. We claim that this direction is correct and that DSGE models should be extensively used in forecasting, even though the detected deficiencies suggest that there may be still some potential for improving their performance. Furthermore, we propose a method of forecast formulation that involves combining experts' nowcasts with DSGE model forecasts for the remaining horizons. We also dissuade from adding expert corrections to forecasts generated using this method.

References

- Adjemian, Stéphane, Matthieu Darracq Paries, and Stéphane Moyen. (2008) "Towards a Monetary Policy Evaluation Framework." Working Paper Series 942, European Central Bank.
- Adolfson, Malin, Jesper Lindé, and Mattias Villani. (2007) "Forecasting Performance of an Open Economy DSGE Model." *Econometric Reviews*, 26, 289–328.
- Berkowitz, Jeremy. (2001) "Testing Density Forecasts, with Applications to Risk Management." *Journal of Business & Economic Statistics*, 19, 465–74.
- Chari, V.V., Patrick J. Kehoe, and Ellen R. McGrattan. (2008) "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?" *Journal of Monetary Economics*, 55, 1337–52.
- Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. (2005) "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy." *Journal of Political Economy*, 113, 1–45.

- Clements, Michael and David Hendry. (1998) *Forecasting Economic Time Series*, Cambridge University Press.
- Croushore, Dean. (2010) “An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data.” *The B.E. Journal of Macroeconomics*, 10, Article 10.
- Croushore, Dean, and Tom Stark. (2001) “A Real-Time Data Set for Macroeconomists.” *Journal of Econometrics*, 105, 111–30.
- Del Negro, Marco, and Frank Schorfheide. (2004) “Priors from General Equilibrium Models for VARs.” *International Economic Review*, 45, 643–73.
- Del Negro, Marco, and Frank Schorfheide. (2009) “Monetary Policy Analysis with Potentially Misspecified Models.” *American Economic Review*, 99, 1415–50.
- Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters. (2007) “On the Fit of New Keynesian Models.” *Journal of Business & Economic Statistics*, 25, 123–43.
- Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay. (1998) “Evaluating Density Forecasts with Applications to Financial Risk Management.” *International Economic Review*, 39, 863–83.
- Edge, Rochelle M., and Refet S. Gurkaynak. (2010) “How Useful are Estimated DSGE Model Forecasts for Central Bankers?” *Brookings Papers on Economic Activity*, 41, 209–59.
- Edge, Rochelle M., Michael T. Kiley, and Jean-Philippe Laforte. (2010) “A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model.” *Journal of Applied Econometrics*, 25, 720–54.
- Eichenbaum, Martin, and Jonas D.M. Fisher. (2007) “Estimating the Frequency of Price Re-Optimization in Calvo-Style Models.” *Journal of Monetary Economics*, 54, 2032–47.
- Fernández-Villaverde, Jesús, Juan F. Rubio-Ramírez, Thomas J. Sargent, and Mark W. Watson. (2007) “ABCs (and Ds) of Understanding VARs.” *American Economic Review*, 97, 1021–26.
- Geraats, Petra M. (2009) “Trends in Monetary Policy Transparency.” *International Finance*, 12, 235–68.

- Guerron-Quintana, Pablo A. (2010) “What You Match Does Matter: The Effects of Data on DSGE Estimation.” *Journal of Applied Econometrics*, 25, 774–804.
- Harvey, David, Stephen Leybourne, and Paul Newbold. (1997) “Testing the Equality of Prediction Mean Squared Errors.” *International Journal of Forecasting*, 13, 281–91.
- Herbst, Edward, and Frank Schorfheide. (2011) “Evaluating DSGE Model Forecasts of Comovements.” Working Paper Series No. 9, Federal Reserve Bank of Philadelphia.
- Kimball, Miles S. (1995) “The Quantitative Analytics of the Basic Neomonetarist Model.” *Journal of Money, Credit and Banking*, 27, 1241–77.
- Lees, Kirdan, Troy Matheson, and Christie Smith. (2007) “Open Economy Forecasting with a DSGE-VAR: Head to Head with the RBNZ Published Forecasts.” *International Journal of Forecasting*, 27, 512–28.
- Newey, Whitney K., and Kenneth D. West. (1987) “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix.” *Econometrica*, 55, 703–08.
- Newey, Whitney K., and Kenneth D. West. (1994) “Automatic Lag Selection in Covariance Matrix Estimation.” *Review of Economic Studies*, 61, 631–53.
- Rosenblatt, Murray. (1952) “Remarks on a Multivariate Transformation.” *Annals Mathematical Statistics*, 23, 470–72.
- Rubaszek, Michał, and Paweł Skrzypczyński. (2008) “On the Forecasting Performance of a Small-Scale DSGE Model.” *International Journal of Forecasting*, 24, 498–512.
- Schorfheide, Frank, Keith Sill, and Maxym Kryshko. (2010) “DSGE Model-Based Forecasting of Non-Modelled Variables.” *International Journal of Forecasting*, 26, 348–73.
- Smets, Frank, and Rafael Wouters. (2003) “An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area.” *Journal of the European Economic Association*, 1, 1123–75.
- Smets, Frank, and Rafael Wouters. (2007) “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review*, 97, 586–606.
- Taylor, John B., and Volker Wieland. (2009) “Surprising Comparative Properties of Monetary Models: Results from a New Data Base.” Working Papers No. 14849, National Bureau of Economic Research.

Table 1: Prior distribution and recursive estimates for model parameters

		Prior distribution			Recursive estimates (posterior median)		
		Type	Mean	StDev.	Min	Av.	Max
Estimated parameters							
investment adj. cost	φ	normal	4.00	1.50	2.27	2.66	3.21
consumption elasticity	σ_c	normal	1.50	0.37	1.27	1.46	1.57
consumption habit	h	beta	0.70	0.10	0.45	0.50	0.58
Calvo wages	ξ_w	beta	0.50	0.10	0.57	0.61	0.62
labor supply elasticity	σ_l	normal	2.00	0.75	0.45	0.55	0.68
Calvo prices	ξ_p	beta	0.50	0.10	0.56	0.59	0.63
indexation wages	ι_w	beta	0.50	0.15	0.49	0.51	0.53
indexation prices	ι_p	beta	0.50	0.15	0.19	0.33	0.39
capital util. adj. cost	ψ	beta	0.50	0.15	0.21	0.29	0.35
price mark-up	λ_p	normal	1.25	0.12	1.77	1.84	1.89
int. rate inflation	r_π	normal	1.50	0.25	1.77	1.87	1.96
int. rate smoothing	ρ	beta	0.75	0.10	0.80	0.82	0.84
int. rate output	r_y	normal	0.12	0.05	0.09	0.10	0.11
int. rate output growth	$r_{\Delta y}$	normal	0.12	0.05	0.22	0.23	0.24
steady-state inflation	$\bar{\pi}$	gamma	0.62	0.10	0.64	0.65	0.66
discount factor	$100(\beta^{-1} - 1)$	gamma	0.25	0.10	0.13	0.15	0.18
steady-state hours	\bar{l}	normal	0.00	2.00	1.18	1.78	2.10
trend output p.c. growth	$100(\gamma - 1)$	normal	0.40	0.10	0.27	0.30	0.33
capital share	α	normal	0.30	0.05	0.15	0.16	0.17
Calibrated parameters							
depreciation rate	δ		0.025	0	0.025	0.025	0.025
gov. spending share	g_y		0.18	0	0.18	0.18	0.18
wage mark-up	λ_w		1.50	0	1.50	1.50	1.50
price elasticity curv.	ε_p		10	0	10	10	10
wage elasticity curv.	ε_w		10	0	10	10	10
trend population growth	$100n$		0.37	0	0.37	0.37	0.37
Shock processes							
productivity	ρ_a	beta	0.50	0.20	0.99	0.99	1.00
	σ_a	inv. gam.	0.10	2.00	0.38	0.40	0.41
risk premium	ρ_b	beta	0.50	0.20	0.47	0.56	0.65
	σ_b	inv. gam.	0.10	2.00	0.62	0.78	0.94
investment	ρ_i	beta	0.50	0.20	0.54	0.59	0.62
	σ_i	inv. gam.	0.10	2.00	0.56	0.59	0.62
price mark-up	ρ_p	beta	0.50	0.20	0.98	0.98	0.99
	μ_p	beta	0.50	0.20	0.81	0.84	0.88
	σ_p	inv. gam.	0.10	2.00	0.12	0.14	0.20
monetary	ρ_r	beta	0.50	0.20	0.15	0.23	0.25
	σ_r	inv. gam.	0.10	2.00	0.21	0.23	0.25
gov. spending	ρ_g	beta	0.50	0.20	0.93	0.95	0.96
	σ_g	inv. gam.	0.10	2.00	0.50	0.52	0.54
	ρ_{ga}	beta	0.50	0.20	0.61	0.68	0.74

Notes: For the inverse gamma distribution, the mode and the degrees of freedom are reported.

Table 2: Recursive estimates of posterior median for the DSGE-VAR weight parameter

	Prior distribution			Recursive estimates (posterior median)			DSGE weight (in %)		
	Type	Min	Max	Min	Av.	Max	Min	Av.	Max
DSGE-VAR2($\hat{\lambda}$)	unif.	0	10	0.44	0.58	0.69	30.8	36.4	40.7
DSGE-VAR4($\hat{\lambda}$)	unif.	0	10	0.71	0.91	1.13	41.5	47.6	53.0
DSGE-VAR6($\hat{\lambda}$)	unif.	0	10	1.14	1.41	1.67	53.2	58.3	62.5
DSGE-VAR8($\hat{\lambda}$)	unif.	0	10	1.61	1.99	2.42	61.7	66.4	70.8

Table 3: Mean Forecast Errors (MFEs) of unconditional forecasts

h	DSGE	SPF	DSGE-VAR($\hat{\lambda}$)				DSGE- VAR(∞) $p = 16$
			$p = 2$	$p = 4$	$p = 6$	$p = 8$	
<i>Output growth (real GDP, QoQ SAAR)</i>							
0	-0.57**	0.38	-0.98***	-0.98***	-0.86***	-0.79***	-0.54*
1	-0.24	0.17	-1.01***	-0.98***	-0.87**	-0.70**	-0.39
2	-0.03	0.12	-1.05***	-0.94**	-0.74*	-0.55	-0.31
3	0.11	0.02	-1.07***	-0.86**	-0.66*	-0.45	-0.28
4	0.07	-0.19	-1.18***	-0.88**	-0.72*	-0.51	-0.37
<i>Inflation (GDP price index, QoQ SAAR)</i>							
0	0.15	0.12	0.14	0.16	0.11	0.08	0.18
1	0.06	0.09	0.05	0.08	0.00	-0.06	0.10
2	-0.02	0.01	-0.06	-0.01	-0.10	-0.17	0.03
3	-0.04	-0.01	-0.14	-0.07	-0.18	-0.25	0.01
4	-0.11	-0.06	-0.26	-0.20	-0.34	-0.40	-0.06
<i>Interest rate (three-month TB rate, per annum)</i>							
0	-0.11	-0.07***	-0.03	-0.05	-0.12	-0.12	-0.10
1	-0.25	-0.19**	-0.14	-0.17	-0.30*	-0.29*	-0.21
2	-0.39	-0.33*	-0.30	-0.36	-0.52**	-0.50**	-0.35
3	-0.51*	-0.47*	-0.47	-0.55*	-0.75**	-0.72**	-0.48
4	-0.65*	-0.65*	-0.67*	-0.75**	-0.98***	-0.94***	-0.64

Notes: A positive value indicates that forecasts are on average below the actual values. Asterisks ***, ** and * denote the rejection of the null that the MFE is equal to zero at 1%, 5% and 10% significance level, respectively. Test statistics are corrected for autocorrelation of forecast errors with the Newey and West (1987, 1994) method.

Table 4: Root Mean Squared Forecast Errors (RMSFEs) of unconditional forecasts

h	DSGE	SPF	DSGE-VAR($\hat{\lambda}$)				DSGE-VAR(∞) $p = 16$
			$p = 2$	$p = 4$	$p = 6$	$p = 8$	
<i>Output growth (real GDP, QoQ SAAR)</i>							
0	1.95	0.98	1.10	1.17**	1.15	1.16	1.04
1	1.99	1.07	1.18	1.23**	1.23*	1.23*	1.10
2	1.83	1.21**	1.23*	1.29**	1.26*	1.25*	1.13*
3	1.89	1.19**	1.23**	1.23**	1.21*	1.19*	1.14**
4	2.10	1.16**	1.21**	1.16**	1.17*	1.18*	1.13**
<i>Inflation (GDP price index, QoQ SAAR)</i>							
0	0.96	0.91	1.05**	0.96	0.99	0.97	1.02
1	0.97	0.96	1.07***	0.93*	1.01	1.00	1.02
2	0.88	1.09	1.09***	0.95	1.03	1.04	1.00
3	1.03	1.05	1.11***	0.97	1.00	1.01	0.94
4	1.11	1.02	1.10***	0.96	1.01	1.01	0.96
<i>Interest rate (three-month TB rate, per annum)</i>							
0	0.43	0.34***	0.86	0.93	0.94	0.94	0.98
1	0.80	0.64**	0.91	1.00	1.01	1.02	1.02
2	1.10	0.79	0.95	1.00	1.03	1.05	1.06
3	1.32	0.91	1.01	1.05	1.09	1.11	1.11
4	1.51	1.02	1.07	1.11	1.15	1.16*	1.15*

Notes: For the DSGE model RMSFEs are reported in levels, whereas for the remaining methods they appear as the ratios to the corresponding RMSFE from the DSGE model. Asterisks ***, ** and * denote the rejection of the null of the HLN-DM test, stating that the RMSFE is not significantly different from the corresponding RMSFE from the DSGE model, at 1%, 5% and 10% significance level, respectively.

Table 5: Efficiency test for unconditional forecasts

h	DSGE				SPF				DSGE-VAR4			
	$\hat{\alpha}_0$ ($S_{\hat{\alpha}_0}$)	$\hat{\alpha}_1$ ($S_{\hat{\alpha}_1}$)	R^2	χ^2 <i>prob</i>	$\hat{\alpha}_0$ ($S_{\hat{\alpha}_0}$)	$\hat{\alpha}_1$ ($S_{\hat{\alpha}_1}$)	R^2	χ^2 <i>prob</i>	$\hat{\alpha}_0$ ($S_{\hat{\alpha}_0}$)	$\hat{\alpha}_1$ ($S_{\hat{\alpha}_1}$)	R^2	χ^2 <i>prob</i>
<i>Output growth (real GDP, QoQ SAAR)</i>												
0	0.59 (0.88)	0.68 (0.21)	0.17	10.3 <i>0.01</i>	1.09 (0.61)	0.74 (0.19)	0.15	3.29 <i>0.19</i>	1.49 (0.94)	0.39 (0.19)	0.05	39.9 <i>0.00</i>
2	0.52 (0.77)	0.82 (0.27)	0.17	0.47 <i>0.79</i>	5.03 (1.17)	-0.73 (0.42)	0.05	18.6 <i>0.00</i>	2.73 (1.31)	0.06 (0.31)	0.00	20.0 <i>0.00</i>
4	0.12 (1.14)	0.98 (0.43)	0.11	0.06 <i>0.97</i>	5.04 (1.55)	-0.78 (0.49)	0.05	15.2 <i>0.00</i>	2.94 (2.62)	-0.05 (0.67)	0.00	13.2 <i>0.00</i>
<i>Inflation (GDP price index, QoQ SAAR)</i>												
0	1.51 (0.40)	0.34 (0.17)	0.08	15.0 <i>0.00</i>	1.22 (0.57)	0.47 (0.26)	0.07	4.58 <i>0.10</i>	1.34 (0.32)	0.42 (0.14)	0.12	18.5 <i>0.00</i>
2	1.16 (0.50)	0.47 (0.22)	0.08	5.83 <i>0.05</i>	2.43 (0.75)	-0.10 (0.32)	0.00	12.1 <i>0.00</i>	0.93 (0.40)	0.57 (0.18)	0.16	5.99 <i>0.05</i>
4	3.01 (0.61)	-0.36 (0.26)	0.03	27.9 <i>0.00</i>	3.91 (0.71)	-0.76 (0.30)	0.10	34.8 <i>0.00</i>	1.76 (0.47)	0.18 (0.21)	0.01	15.9 <i>0.00</i>
<i>Interest rate (three-month TB rate, per annum)</i>												
0	-0.75 (0.10)	1.16 (0.03)	0.95	53.8 <i>0.00</i>	0.00 (0.03)	0.98 (0.01)	0.99	11.4 <i>0.00</i>	-0.21 (0.15)	1.04 (0.03)	0.94	2.03 <i>0.36</i>
2	-2.34 (0.63)	1.46 (0.17)	0.66	21.0 <i>0.00</i>	0.00 (0.30)	0.92 (0.07)	0.76	4.17 <i>0.12</i>	-0.43 (0.66)	1.02 (0.16)	0.58	2.36 <i>0.31</i>
4	-3.15 (1.42)	1.57 (0.33)	0.41	9.14 <i>0.01</i>	0.22 (0.87)	0.80 (0.19)	0.34	4.71 <i>0.09</i>	-0.08 (1.50)	0.85 (0.34)	0.23	4.29 <i>0.12</i>

Notes: χ^2 statistics relate to the null of the forecast unbiasedness test ($\alpha_0 = 0$ and $\alpha_1 = 1$). All statistics are corrected for heteroskedasticity and autocorrelation of the residuals with the Newey and West (1987, 1994) method.

Table 6: RMSFEs of forecasts conditional on SPF nowcasts

h	DSGE	SPF	DSGE-VAR($\hat{\lambda}$)				DSGE-VAR(∞) $p = 16$
			$p = 2$	$p = 4$	$p = 6$	$p = 8$	
<i>Output growth (real GDP, QoQ SAAR)</i>							
1	1.93	1.10	1.11	1.18*	1.17*	1.20**	1.08
2	1.87	1.18*	1.16	1.22*	1.22*	1.23*	1.13*
3	1.82	1.23***	1.26**	1.29**	1.29**	1.27*	1.20***
4	2.11	1.16**	1.19**	1.13*	1.16*	1.18*	1.14**
<i>Inflation (GDP price index, QoQ SAAR)</i>							
1	0.92	1.01	1.09***	0.97	1.03	1.01	1.02
2	0.92	1.05	1.07**	0.96	1.02	1.01	0.98
3	1.01	1.06*	1.09***	0.97	0.98	0.99	0.96
4	1.11	1.02	1.10***	0.96	0.98	0.98	0.95
<i>Interest rate (three-month TB rate, per annum)</i>							
1	0.55	0.93	0.96	1.00	1.01	1.01	1.02
2	0.92	0.94	0.95	0.99	1.02	1.03	1.03
3	1.21	0.99	0.98	1.02	1.04	1.06	1.07
4	1.45	1.07	1.02	1.05	1.08	1.09	1.11

Notes: See notes to Table 4.

Table 7: RMSFEs conditional on interest rate SPF nowcast

h	DSGE	SPF	DSGE-VAR				DSGE-VAR(∞) $p = 16$
			$p = 2$	$p = 4$	$p = 6$	$p = 8$	
<i>Output growth (real GDP, QoQ SAAR)</i>							
0	2.33	0.82	1.00	1.07	1.07	1.10	1.02
1	2.05	1.04	1.18	1.23*	1.23**	1.23**	1.11*
2	1.84	1.20**	1.24*	1.31*	1.28*	1.27*	1.14*
3	1.89	1.19**	1.23**	1.22*	1.21*	1.19*	1.14**
4	2.12	1.15**	1.20**	1.15*	1.17*	1.18*	1.11*
<i>Inflation (GDP price index, QoQ SAAR)</i>							
0	0.94	0.93	1.06*	0.97	1.01	0.98	1.03
1	0.96	0.97	1.08*	0.94	1.02	1.00	1.03
2	0.85	1.12	1.12*	0.97	1.06	1.07	1.03
3	1.00	1.08	1.14***	1.00	1.04	1.04	0.96
4	1.08	1.05	1.12***	0.98	1.04	1.04	0.98
<i>Interest rate (three-month TB rate, per annum)</i>							
1	0.60	0.86	0.97	1.02	1.07	1.07	1.07
2	0.99	0.87	0.96	1.03	1.08	1.09	1.08
3	1.28	0.94	1.00	1.04	1.09	1.10	1.11
4	1.50	1.03	1.05	1.08	1.13	1.13	1.15

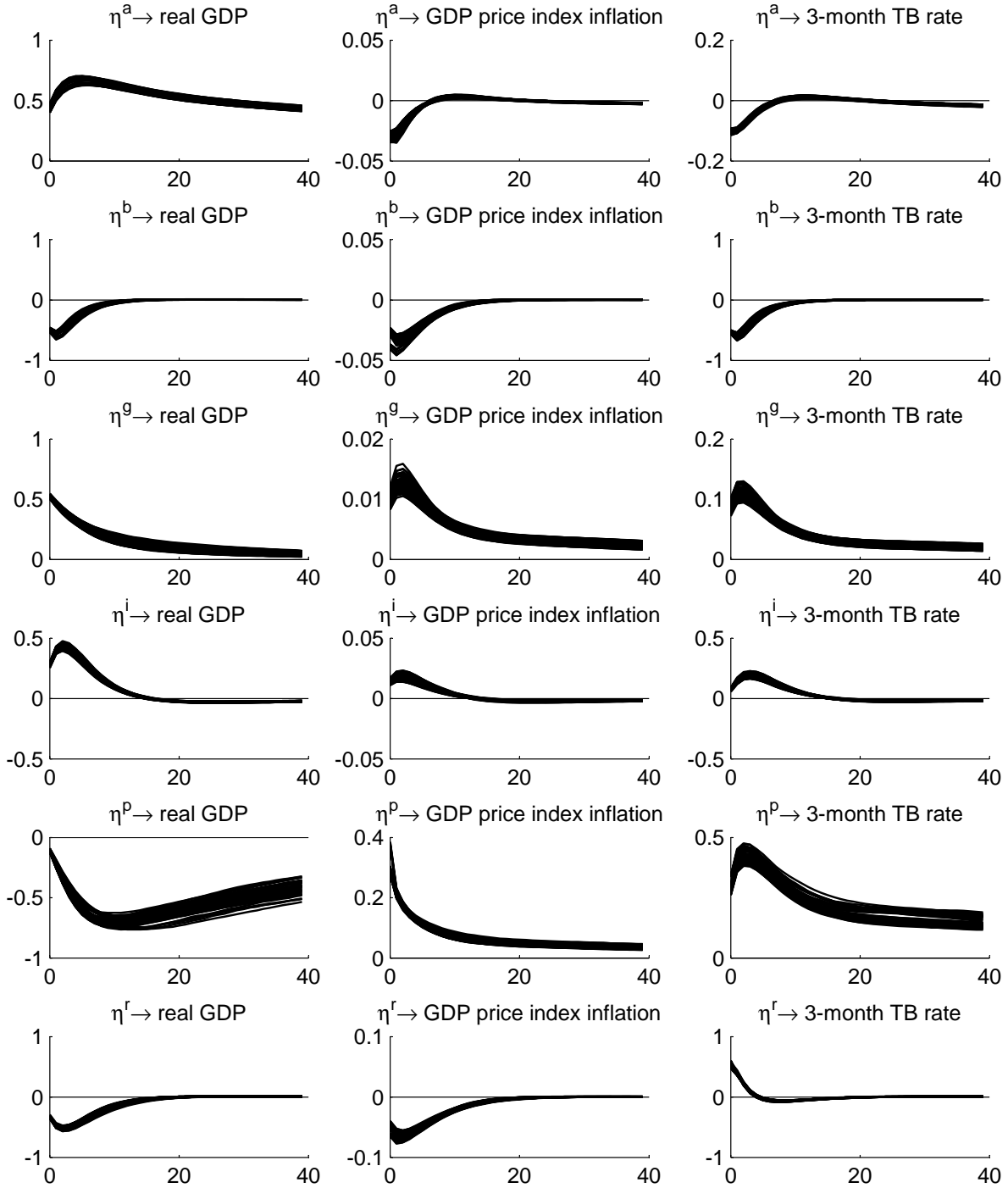
Notes: See notes to Table 4.

Table 8: Berkowitz test for uniform and independent distribution of PITs

h	DSGE				DSGE-VAR4			
	$\hat{\gamma}_0$ ($S_{\hat{\gamma}_0}$)	$\hat{\gamma}_1$ ($S_{\hat{\gamma}_1}$)	$\hat{\sigma}_\eta$ $\hat{\sigma}_z$	LR $prob$	$\hat{\gamma}_0$ ($S_{\hat{\gamma}_0}$)	$\hat{\gamma}_1$ ($S_{\hat{\gamma}_1}$)	$\hat{\sigma}_\eta$ $\hat{\sigma}_z$	LR $prob$
	<i>Output growth (real GDP, QoQ SAAR)</i>							
0	-0.15 (0.07)	-0.08 (0.10)	0.49 0.49	37.2 <i>0.00</i>	-0.34 (0.12)	-0.02 (0.12)	0.71 0.71	16.9 <i>0.00</i>
2	0.00 (0.06)	0.03 (0.16)	0.45 0.45	41.9 <i>0.00</i>	-0.23 (0.11)	0.19 (0.15)	0.66 0.67	20.5 <i>0.00</i>
4	0.02 (0.07)	0.10 (0.16)	0.52 0.52	30.4 <i>0.00</i>	-0.20 (0.10)	0.27 (0.15)	0.66 0.69	19.9 <i>0.00</i>
	<i>Inflation (GDP price index, QoQ SAAR)</i>							
0	0.13 (0.11)	0.01 (0.11)	0.80 0.80	5.53 <i>0.14</i>	0.15 (0.12)	0.07 (0.12)	0.88 0.88	3.36 <i>0.34</i>
2	0.00 (0.06)	0.51 (0.10)	0.46 0.53	46.3 <i>0.00</i>	0.00 (0.07)	0.45 (0.10)	0.55 0.61	31.9 <i>0.00</i>
4	-0.04 (0.08)	0.36 (0.18)	0.56 0.60	28.8 <i>0.00</i>	-0.09 (0.10)	0.23 (0.18)	0.64 0.66	18.5 <i>0.00</i>
	<i>Interest rate (three-month TB rate, per annum)</i>							
0	-0.08 (0.07)	0.62 (0.10)	0.52 0.67	41.3 <i>0.00</i>	-0.06 (0.07)	0.49 (0.12)	0.51 0.58	36.8 <i>0.00</i>
2	-0.07 (0.07)	0.89 (0.06)	0.42 0.92	84.4 <i>0.00</i>	-0.08 (0.07)	0.86 (0.08)	0.44 0.87	73.5 <i>0.00</i>
4	-0.07 (0.07)	0.94 (0.06)	0.37 1.06	107.1 <i>0.00</i>	-0.09 (0.07)	0.93 (0.06)	0.39 1.04	103.8 <i>0.00</i>

Notes: LR statistics relate to the null of the Berkowitz test. All statistics are corrected for heteroskedasticity and autocorrelation of the residuals with the Newey and West (1987, 1994) method.

Figure 1: Recursive impulse response functions



Notes: Each line corresponds to a different quarter of the evaluation sample.

Figure 2: Actuals and four-quarter-ahead forecasts

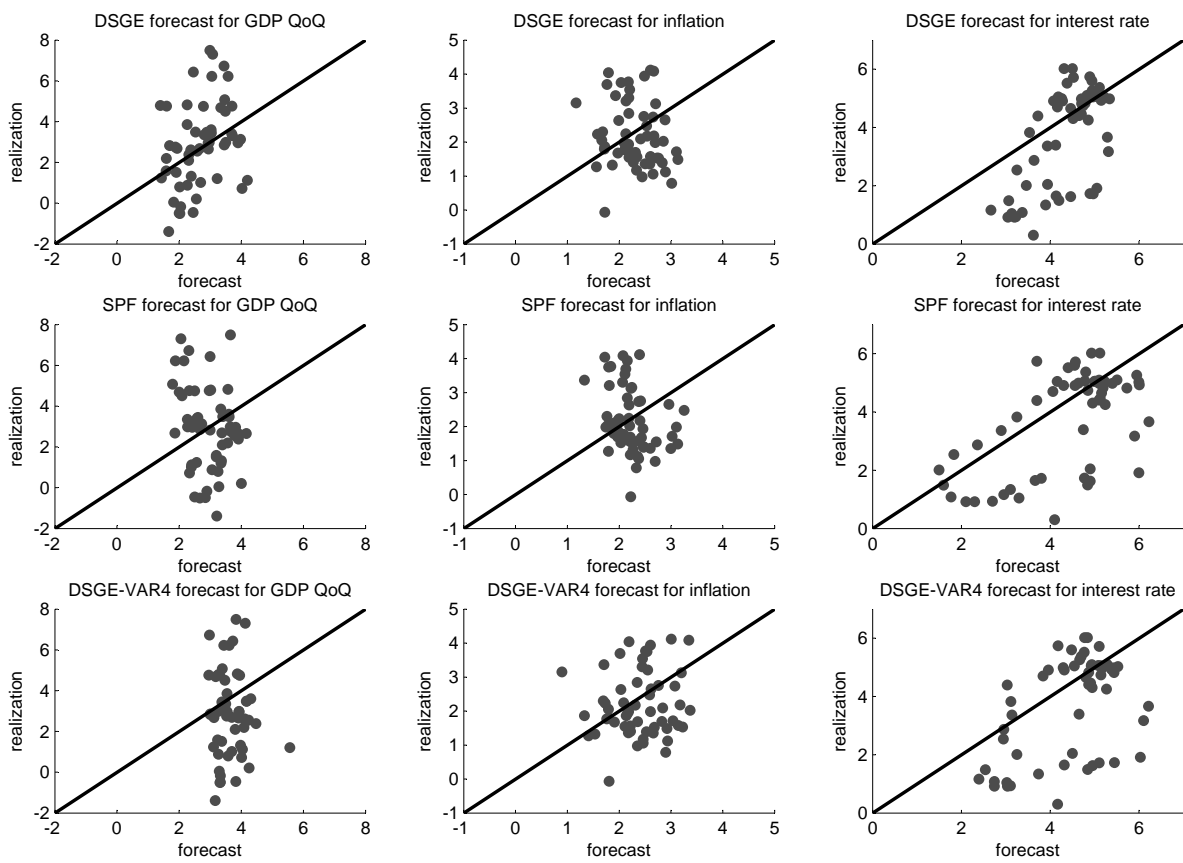
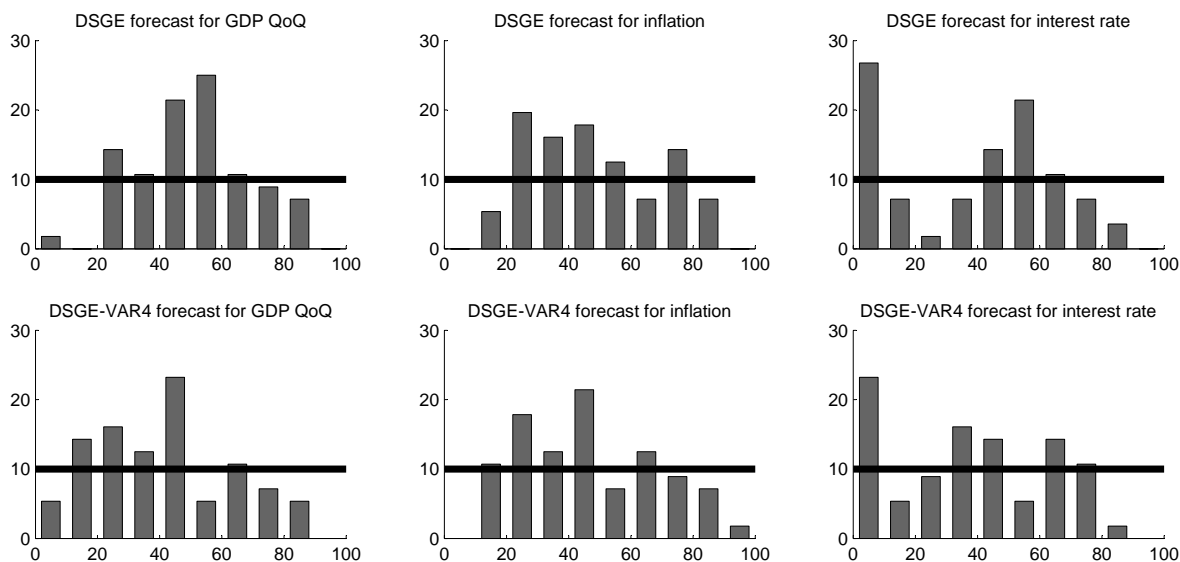


Figure 3: Density forecasts: PIT histograms for four-quarter-ahead forecasts



Notes: Bars represent the fraction of realized observations falling into deciles of density forecasts. The theoretical value of 10% for a well calibrated model is represented by a solid line.