



Department of Applied Econometrics Working Papers

Warsaw School of Economics
Al. Niepodległości 164
02-554 Warszawa, Poland

Working Paper No. 5-07

Can a simple DSGE model outperform Professional Forecasters?

Michał Rubaszek, Paweł Skrzypczyński
National Bank of Poland and Warsaw School of Economics

This paper is available at the Warsaw School of Economics
Department of Applied Econometrics website at: <http://www.sgh.waw.pl/instytut/zes/wp/>

This paper is also published at National Bank of Poland website at:
http://www.nbp.pl/publikacje/materialy_i_studia/43_en.pdf

Can a simple DSGE model outperform Professional Forecasters?

Michał Rubaszek^{*}
Paweł Skrzypczyński^{**}

September, 2007

Abstract

DSGE models have recently become one of the most frequently used tools in policy analysis. Nevertheless, their forecasting proprieties are still unexplored. In this article we address this problem by examining the quality of forecasts from a small size DSGE model, a trivariate VAR model and the Philadelphia Fed Survey of Professional Forecasters. The forecast performance of these methods is analysed for the key U.S. economic variables: the three month Treasury bill yield, the GDP growth rate and the GDP price index inflation. We evaluate the *ex post* forecast errors on the basis of the data from the period of 1994-2006. We apply the Philadelphia Fed “Real-Time Data Set for Macroeconomists,” described by Croushore and Stark (2001a), to ensure that the information available to the SPF was exactly the same as the data used to estimate the DSGE and VAR models.

Overall, the results are mixed. It appears that when comparing the root mean squared errors for some forecast horizons the DSGE model seems to outperform the SPF in forecasting the GDP growth rate. However, this characteristic turned out to be not statistically significant. In principle most forecasts of the GDP price index inflation and the short term interest rate by the SPF are significantly better than those from the DSGE model. The forecast quality of the VAR model turned out to be the worst one.

Keywords: forecasting; real-time data; Survey of Professional Forecasters; DSGE; VAR.

JEL Classification: C32, C53, E12, E17.

^{*} National Bank of Poland, Warsaw School of Economics, e-mail: Michal.Rubaszek@mail.nbp.pl.

^{**} National Bank of Poland, e-mail: Pawel.Skrzypczynski@mail.nbp.pl.

Introduction

Forecasting inflation, output and interest rates in the United States is one of the crucial tasks for many domestic and foreign financial institutions and other economic entities. The reason is that the ability to predict future state of the U.S. economy accurately facilitates the decision-taking process. For instance, the precise forecast of short term interest rates would be the useful information for an investment fund while setting duration of its bond portfolio. Similarly, many central banks would like to know more about future economic situation in the United States, while setting the level of domestic interest rates. As a result the question arises, which method is the most appropriate in forecasting the U.S. economy. We explore this issue by comparing the forecast performance of a small-scale dynamic stochastic general equilibrium (DSGE) model, the Survey of Professional Forecasters (SPF), as well as a vector autoregression (VAR) model.

DSGE models have recently become one of the most frequently used tools for quantitative policy analysis in macroeconomics, mainly because of their characteristics such as micro-foundations or explicitly modeled expectations. Nevertheless, as stated by Smets and Wouters (2004), these models are hardly applied for forecasting, on the basis of the argument that they perform poorly in this field. Del Negro et al (2005) claim that due to the improved time series fit of these models, their role in forecasting should increase. Although few central banks have recently decided to use DSGE models for inflation projections, the discussion about the application of DSGE models in macroeconomic forecasting is still open. Moreover, the documentation of the DSGE models out-of-sample performance is still scarce. We think that this issue is of special importance as the growing use of DSGE models requires an answer to the question about their abilities to forecast future state of the economy.

We believe that the SPF, which is the oldest regularly carried out survey of macroeconomic forecasts, represents plausible approximation of the market expectations about the future economic situation in the United States. Therefore we regard that the SPF represents the best reference point in evaluating the accuracy of the forecasts from the DSGE model.

Since the publication of Sims (1980), the small-scale VAR models have been widely applied in macroeconomics, both for policy analysis and forecasting. It should be noted that infinite order VARs constitute unconstrained versions of DSGE models, and hence these a-theoretical models have been widely used as a benchmark for evaluating the performance of DSGE models by comparing the impulse-response functions, as in Christiano et al (2005), or

forecast errors, as in Smets and Wouters (2004). Better forecast accuracy of the DSGE model than that of the VAR model would therefore justify constraints given by the economic theory. Bearing this in mind, we also investigate whether the quality of forecasts from the DSGE model and the SPF are superior to that from the VAR model.

As stated by Croushore and Stark (2001a), while comparing the *ex post* forecast performance of an estimated model with the SPF on the basis of the latest-available data, the researcher is favouring his model for the following two reasons. Firstly, he knows the *ex post* realizations of the data and hence has richer data set for building a model. Secondly, the latest-available data may substantially differ from those disposable to the SPF due to data revisions. We bear in mind that the forecasters could not use estimated DSGE models in the mid 90s, since this kinds of models were not well developed then. Consequently, we are not able to eliminate the first advantage. We address the second favour by applying the Philadelphia Fed “Real-Time Data Set for Macroeconomists,” described by Croushore and Stark (2001a), to our analysis. Therefore, the information, which was available to the SPF, is exactly the same as the data applied to estimate the DSGE and VAR models.

As discussed in Croushore (2006), evaluating the accuracy of real-time forecasts requires taking a decision on what to consider as the actual data in calculating forecast errors. We tackle this problem in two ways. First, we analyse the forecast performance by taking the latest available data set, i.e. from the vintage of the third quarter of 2006, as the realisation of the variables. In the second case we compare the forecasts with real-time data available one year after a given date of the vintage used in estimation.

Overall, the results are mixed. It appears that when comparing the root mean squared errors for some forecast horizons the DSGE model seems to outperform the SPF in forecasting the GDP growth rate. However, this characteristic turned out to be not statistically significant. In principle most forecasts of the GDP price index inflation and the short term interest rate by the SPF are significantly better than those from the DSGE model. The forecast quality of the VAR model turned out to be the worst one.

The contribution of the paper is twofold. Firstly, we extend the knowledge about the forecasting properties of small-scale DSGE models. Secondly, we believe that this is the first study that compares the forecast errors from a DSGE model with those from the SPF in a real-time environment. Our results confirm that the importance of DSGE models in forecasting should increase.

The paper is organized as follows. In section 1 we examine the literature that discusses the forecasting performance of the SPF, DSGE and VAR models, both in real-time and latest-

available data contexts. Section 2 introduces three models applied to generate forecasts: the SPF, the DSGE and the VAR. In section 3 we describe real-time data set used in the estimation of the DSGE and VAR models. Section 4 presents the results of the out-of-sample forecast performance analysis. We focus there on the *ex post* forecast errors from the models described in the previous section. The forecast accuracy is evaluated on the basis of the data from the period of 1994-2006. We conclude in the last section.

1. Literature review

The number of articles that evaluate the forecasting proprieties of DSGE models in a real-time environment is scarce. According to our best knowledge, the only such analysis has been elaborated by Edge, Kiley and Laforge (2006).¹ The authors compare mean absolute errors (MAE) of forecasts from the random walk, a VAR, a BVAR and a richly-specified DSGE model to the Federal Reserves staff projections. The variables considered are GDP growth rate, real consumption growth rate, GDP price index and PCE inflation rates in the United States. The authors find that the FRB staff is the best inflation forecaster, while the DSGE, VAR and BVAR models dominate in forecasting the GDP growth rate. It should be noted that since the forecast performance evaluation period of 1996-2000 is short, the results might not be representative.

There are few articles that compare the quality of the forecasts from DSGE to those from VAR models. However, since these analyses are not carried out in a real-time context, the results might be biased. The two most notable examples are papers by Smets and Wouters (2004) and Del Negro et al (2005). The former article illustrates how a medium-scale DSGE model for the euro area, described by Smets and Wouters (2003), can be applied for macroeconomic projections and economic analysis. The authors compare the root mean squared errors (RMSE) of the forecasts from the DSGE model to those from VARs for seven macroeconomic variables, among them output growth, inflation and nominal interest rates. According to the results, which might be not representative since the forecast evaluation sample consists of only 16 quarterly observations from 1999-2002, the DSGE moderately outperforms the VAR models. The latter article develops a DSGE-VAR model, which can be characterized as a BVAR model with priors deriving from a DSGE model. The authors apply this new concept to the previously mentioned Smets and Wouters (2003) DSGE framework. Subsequently, on the basis of the rolling sample from 1985-2000, they compare forecast

¹ It should be noted that their article is a preliminary version and has not been published yet.

errors for seven key variables describing the U.S. economy. As stated by the authors, the RMSEs of the forecasts from the DSGE and VAR models are generally comparable, but the forecast accuracy of both these models appear to be inferior to this of the DSGE-VAR model.

A number of papers compare the quality of real-time forecasts from the SPF to univariate a-theoretical models like ARMA or to VAR ones. For example, Croushore (2006) analyses the SPF forecast errors for the U.S. GDP price index inflation over the period 1971-2004. The author tests and accepts the hypothesis that the SPF forecasts are unbiased, except for the sub-sample period 1971-1981, when the U.S. inflation was affected by the oil-price shocks. Subsequently, he compares forecast errors from the SPF and a simple ARIMA model estimated on the basis of real-time data. According to the results, there is little evidence that this univariate model can outperform the Professional Forecasters. As stated by the author, the best evidence in favour of ARIMA models in the earlier studies is derived from using latest available data rather than the real-time ones. Clark and McCracken (2006) present more extensive study comparing the real-time forecast accuracy of the SPF, the Federal Reserve staff and numerous a-theoretical models. The results indicate that the performance of VAR and univariate models in forecasting the GDP growth rate, the GDP price index inflation, the CPI inflation and the Treasury bill rate is roughly the same. The forecast errors from the VARs are, however, significantly higher than those from the SPF, especially for a one-quarter forecast horizon. Finally, the SPF appeared to be more successful than the FRB staff in forecasting the GDP growth rate, but less successful in case of the GDP price index and CPI inflation.

The general picture that emerges from the above studies is that in a real-time context the SPF can better forecast the economy than a-theoretical models like VARs or ARIMAs. Furthermore, if the forecast performance is evaluated on the basis of the latest available data, it appears that DSGE models are comparable or even superior to the a-theoretical ones. The question arises, whether DSGE models can beat the SPF in forecasting the U.S. economy if real-time data are used. Providing an answer is the main purpose of this article.

2. The models

In this section we present three methods that are subsequently applied to forecast key macroeconomic variables of the U.S. economy. We start with an extensive description of the structure of a small scale DSGE model, which may be classified as the microfounded, forward-looking, New Keynesian model. We also discuss issues concerning estimation of

such kind of models. Then, we focus on a trivariate VAR for output, inflation and short term interest rates. Finally, we give a brief outline of the Survey of Professional Forecasters.

2.1. Dynamic Stochastic General Equilibrium model

The model economy is populated by three groups of agents: households that optimize their lifetime utility, firms that maximize profits and monetary authorities that, according to the law, care for price and output stability. A log-linearized version of the model consists of three core key equations: a dynamic IS curve, a forward looking Phillips curve and a monetary policy rule, which determine the path of output, prices and short term nominal interest rates. The system is put in motion by three structural shocks. The first one, productivity shock, affects the level of production technology. In comparison to the real business cycle model of Kydland and Prescott (1982), we assume that productivity is an I(1) process. The second, demand shock, impacts households' decisions concerning consumption and savings. The third, monetary shock derives from monetary authorities' decisions.

2.1.1. Firms

Production of consumption good in the model economy is divided into two stages. In the first stage, firms indexed by $k \in [0,1]$ operating at a monopolistically competitive market are producing differentiated intermediate goods (Y_t^k) which are sold at price (P_t^k) to producers of final good. In the second stage, intermediate goods are transformed into homogenous final good by perfectly competitive firms. They assemble the final good using a constant returns to scale technology of Dixit and Stiglitz (1977):

$$Y_t = \left[\int_0^1 (Y_t^k)^{\frac{\theta-1}{\theta}} dk \right]^{\frac{\theta}{\theta-1}}, \quad (1)$$

where $\theta > 1$ is elasticity of intra-temporal substitution. Final good producers minimize the cost of elaborating their output Y_t by deciding on the amount of each differentiated intermediate good Y_t^k they purchase, taking P_t^k as given. The minimal cost is hence equal to:

$$P_t = \left[\int_0^1 (P_t^k)^{1-\theta} dk \right]^{\frac{1}{1-\theta}} \quad (2)$$

and constitutes the price of final good, which is sold to the consumers. The optimal decision of final good producers also determines the demand for k -th intermediate good as:

$$Y_t^k = \left(\frac{P_t^k}{P_t} \right)^{-\theta} Y_t. \quad (3)$$

Each differentiated good is produced by one firm that uses L_t^k units of labour as the only input. The total output Y_t^k is given by the production function with constant returns to scale:

$$Y_t^k = A_t L_t^k - \frac{Y_t}{\theta}, \quad (4)$$

where expression Y_t/θ represents fixed costs guarantying that in equilibrium profits are null. Technology A_t is assumed to be a nonstationary process given by:

$$A_t = g A_{t-1} \varepsilon_t^S, \quad (5)$$

where g is growth rate of technology in steady state and ε_t^S is a supply shock following an AR(1) process $\varepsilon_t^S = (1 - \rho^S) \bar{\varepsilon}^S + \rho^S \varepsilon_{t-1}^S + \eta_t^S$ with IID white noise disturbance. The first-order condition for cost minimization implies that the nominal marginal cost per unit of output equals to:

$$MC_t^N = \frac{W_t}{A_t}. \quad (6)$$

The instantaneous profits, which are transferred to households in form of dividends, are thus given by:

$$D_t^k = \left(P_t^k - MC_t^N \right) \left(\frac{P_t^k}{P_t} \right)^{-\theta} Y_t - \frac{P_t Y_t}{\theta}. \quad (7)$$

In the flexible price environment, in each period intermediate goods producers would optimize their profits by setting price P_t^k of their output, taking P_t and Y_t as given. The optimal price would be equal to the mark-up over the marginal cost:

$$P_t^k = \frac{\theta}{\theta-1} MC_t^N. \quad (8)$$

We suppose, however, that firms are not able to set their prices in each period. Instead, we introduce nominal stickiness into the model economy by assuming that prices are set within the staggered contract framework described by Calvo (1983). In each period the representative firm is allowed to set the price of its output at a desirable level with probability $(1 - \xi)$. In other case the price is automatically adjusted by the steady state inflation rate $(\bar{\Pi})$ and a fraction δ of the last period's excessive inflation rate.² Thus, if firm k has not re-optimized the price of its output since period t , then the price in period $t + s$ equals to:

$$P_{t+s}^k = P_t^k \left(\frac{P_{t+s-1}}{P_{t-1} \bar{\Pi}^s} \right)^\delta (\bar{\Pi})^s. \quad (9)$$

Producers that are allowed to re-optimize their price are maximizing the present value of discounted inter-temporal profits:

$$\max_{P_t^k} E_t \left\{ \sum_{s=0}^{\infty} \xi^s Q_{t,t+s} D_{t+s}^k \right\}, \quad (10)$$

where $Q_{t,t+s}$ is stochastic discount factor. Substituting equations (7) and (9) into equation (10) yields the following optimization problem:

$$\max_{P_t^k} E_t \left\{ \sum_{s=0}^{\infty} \xi^s Q_{t,t+s} \left[\left(P_t^k \left(\frac{P_{t+s-1}}{P_{t-1}} \right)^\delta \bar{\Pi}^s - MC_{t+s}^N \right) \left(\frac{P_t^k \left(\frac{P_{t+s-1}}{P_{t-1}} \right)^\delta \bar{\Pi}^s}{P_{t+s}} \right)^{-\theta} - \frac{P_{t+s}}{\theta} Y_{t+s} \right] \right\}. \quad (11)$$

The first-order condition for this maximization problem is:

$$E_t \left\{ \sum_{s=0}^{\infty} \xi^s Q_{t,t+s} Y_{t+s}^k \left(\tilde{P}_t^k \left(\frac{P_{t+s-1}}{P_{t-1}} \right)^\delta \bar{\Pi}^s - \frac{\theta}{\theta-1} MC_{t+s}^N \right) \right\} = 0, \quad (12)$$

² Similar indexation patterns were introduced by Smets and Wouters (2003) or Christiano et al (2005).

where \tilde{P}_t^k is the level of price that maximizes the expected value of future dividends. As a result, according to equation (12), the chosen price positively depends on current and expected future marginal costs.

Finally, according to the definition of the aggregate price index given by equation (2), the price level equals to:

$$P_t = \left[\xi \left(P_{t-1} \left(\frac{P_{t-1}}{P_{t-2}} \right)^\delta \bar{\Pi} \right)^{1-\theta} + (1-\xi) (\tilde{P}_t^k)^{1-\theta} \right]^{\frac{1}{1-\theta}}. \quad (13)$$

2.1.2. Households

The model economy is populated by a continuum of homogenous households indexed by $i \in [0,1]$. In each period t typical household maximizes the lifetime utility function:

$$E_t \sum_{s=0}^{\infty} \beta^{t+s} U_{t+s}^i (H_{t+s}^i; L_{t+s}^i; A_{t+s}; \varepsilon_{t+s}^D), \quad (14)$$

where $\beta < 1$ is time-invariant discount factor. The utility function U_t^i is an increasing function of instantaneous consumption of the representative household with respect to a fraction λ of aggregate past consumption adjusted for the growth rate of technology g , called henceforth habit:

$$H_t^i = C_t^i - \lambda g C_{t-1} \quad (15)$$

and a decreasing function of labour supplied by the typical household L_t^i :

$$U_t^i = E_t \sum_{s=0}^{\infty} \beta^s \varepsilon_{t+s}^D \left\{ \frac{(H_{t+s}^i / A_{t+s})^{1-\sigma}}{1-\sigma} - v_L \frac{(L_{t+s}^i)^{1+\varphi}}{1+\varphi} \right\}. \quad (16)$$

The coefficient σ is the inverse of the intertemporal elasticity of substitution, φ is the inverse of the labour supply elasticity with respect to real wages and ε_t^D is a demand shock that is assumed to be an AR(1) process $\varepsilon_t^D = (1 - \rho^D) \bar{\varepsilon}^D + \rho^D \varepsilon_{t-1}^D + \eta_t^D$, where η_t^D is IID white noise disturbance.

In each period the representative household receives nominal remuneration for the work effort $W_t L_t^i$, dividends from owned firms D_t^i and repayment of funds previously invested in one-period bonds B_{t-1}^i . The money is spent on consumption $P_t C_t^i$ or invested in debt securities, which are sold at a discount rate $1/R_t$. As a result, the budget constraint of the representative household is of the following form:

$$\frac{B_t^i}{R_t} + P_t C_t^i = B_{t-1}^i + W_t L_t^i + D_t^i. \quad (17)$$

In order to maximize the inter-temporal utility function (14) subject to the budget constraint (17) the typical household must make two decisions. First, it must choose how much money should be spent on current consumption and how much should be invested in bonds. The solution of this problem leads to the specification of the dynamic IS curve:

$$\left(\frac{H_t^i}{A_t} \right)^{-\sigma} = \beta E_t \left\{ \frac{\varepsilon_{t+1}^D}{\varepsilon_t^D} \left(\frac{R_t}{\Pi_{t+1}} \right) \frac{A_t}{A_{t+1}} \left(\frac{H_{t+1}^i}{A_{t+1}} \right)^{-\sigma} \right\}, \quad (18)$$

where $\Pi_t = P_t/P_{t-1}$ is inflation rate. Second, the typical household must decide how much time it is eager to spend at work. On the one side, higher labour intensity increases its revenue from remuneration but on the other, it lessens the amount of its leisure time. The outcome of the optimization is the labour supply curve:

$$\frac{W_t}{P_t A_t} = v_L \left(\frac{H_t^i}{A_t} \right)^\sigma (L_t^i)^\varphi. \quad (19)$$

2.1.3. Monetary authorities and market clearing condition

The central bank is supposed to be obliged by law to minimize variation of inflation and output. Short term nominal interest rate is hence adjusted, as in Taylor (1993), in response to deviations of these two variables from their steady-state level. Following Rudebush (2002) and Orphanides and Williams (2002) we extended the original Taylor's specification by introducing variations of output growth and interest smoothing into the monetary policy reaction function:

$$\frac{R_t}{\bar{R}} = \left(\frac{R_{t-1}}{\bar{R}} \right)^\gamma \left[\left(\frac{\Pi_t}{\bar{\Pi}} \right)^{(\gamma_\pi)} \left(\frac{Y_t}{\bar{Y}_t} \right)^{(\gamma_y)} \left(\frac{Y}{Y_{t-1}g} \right)^{(\gamma_{dy})} \right]^{(1-\gamma)} e^{(\eta_t^M)}. \quad (20)$$

Monetary shock η^M is assumed to follow IID white noise process.

The model is closed by specifying clearing condition on goods market. Aggregate supply is equal to aggregate demand if and only if consumption is equal to the production of final good:

$$Y_t = C_t. \quad (21)$$

2.1.4. Steady state and log-linearized version of the model

In symmetric equilibrium all intermediate goods producing firms set their price at the same level ($\bar{P}_t^k = \bar{P}_t$). Consequently, according to equations (3), (7) and (8), output is equal across firms ($\bar{Y}_t^k = \bar{Y}_t$) and profits are null ($\bar{D}_t^k = 0$). As bond holdings across households are the same, their value in the steady-state must satisfy the condition $\bar{B}_t^i = \bar{B}_{t-1}^i = 0$. Finally, budget constraint (17) yields that the total real labour income is spent on consumption ($\bar{C}_t = \frac{\bar{W}_t}{\bar{P}_t} \bar{L}_t$). For the reason that in equilibrium dynamics of some variables depend on the level of technology A_t or price index P_t , we introduce the following de-trended variables $y_t = Y_t/A_t$, $c_t = C_t/A_t$, $h_t = H_t/A_t$, $mc_t = MC_t^N/P_t$ and $w_t = W_t/P_t A_t$. In the absence of structural shocks, the economy converges to a stationary steady-state which is given by \bar{y} , \bar{c} , \bar{h} , \bar{L} , \bar{w} , \bar{mc} , $\bar{\Pi}$ and \bar{R} .

Fluctuations of the economy around this stationary equilibrium are specified by a log-linearized version of the model, in which deviations of a variable x_t from its steady state value are represented by $\hat{x}_t = \log(x_t/\bar{x})$.

Combining definition (15) and the clearing condition (21) yields the relationship between habit, output and supply shock:

$$(1-\lambda)\hat{h}_t = \hat{y}_t - \lambda\hat{y}_{t-1} + \lambda\hat{\varepsilon}_t^S, \quad (22)$$

where the dynamics of habit is given by the Euler equation (18):

$$\hat{h}_t = -\frac{1}{\sigma} \left[\hat{R}_t - \hat{\Pi}_{t+1} - \hat{\varepsilon}_{t+1}^S + \hat{\varepsilon}_{t+1}^D - \hat{\varepsilon}_t^D \right] + \hat{h}_{t+1}. \quad (23)$$

Linearization of the aggregate price index equation (13), taking into account the first order condition for profit maximization defined in (12), leads to the specification of the forward looking Phillips curve:

$$\hat{\Pi}_t = \frac{\delta}{1 + \beta\delta} \hat{\Pi}_{t-1} + \frac{\beta}{1 + \beta\delta} \hat{\Pi}_{t+1} + \frac{(1 - \beta\xi)(1 - \xi)}{(1 + \beta\delta)^\xi} m\hat{c}_t, \quad (24)$$

where, given equations (4), (6) and (19), real marginal costs are equal to:

$$m\hat{c}_t = \sigma \hat{h}_t + \varphi \hat{y}_t. \quad (25)$$

Subsequently, a log-linear approximation of the monetary policy reaction function (20) yields:

$$\hat{R}_t = \gamma \hat{R}_{t-1} + (1 - \gamma) \left(\gamma_\pi \hat{\Pi}_t + \gamma_y \hat{y}_t + \gamma_{\Delta y} (\Delta \hat{y}_t + \hat{\varepsilon}_t^S) \right) + \eta_t^M. \quad (26)$$

The two last equations of the model specify the law of motion for the demand shock:

$$\hat{\varepsilon}_t^D = \rho^D \hat{\varepsilon}_{t-1}^D + \eta_t^D \quad (27)$$

and the supply shock:

$$\hat{\varepsilon}_t^S = \rho^S \hat{\varepsilon}_{t-1}^S + \eta_t^S. \quad (28)$$

2.1.5. Estimation

The system of equations (22)-(28) form a model of unobservable variables that are driven by three structural shocks. This model might be written as:

$$\mathbf{A}_0(\Theta) \mathbf{X}_t = \mathbf{A}_1(\Theta) \mathbf{X}_{t-1} + \mathbf{A}_2(\Theta) \boldsymbol{\eta}_t, \quad (29)$$

where \mathbf{A}_i for $i=0,1,2$ denote matrices that depend on the vector of structural parameters

$\Theta = [\lambda \ \sigma \ \varphi \ \beta \ \xi \ \delta \ \gamma \ \gamma_\pi \ \gamma_y \ \gamma_{\Delta y} \ \rho^S \ \rho^D]'$, \mathbf{X}_t is the vector of model variables $\mathbf{X}_t = [\hat{\Pi}_{t+1} \ \hat{h}_{t+1} \ \hat{\varepsilon}_{t+1}^S \ \hat{\varepsilon}_{t+1}^D \ \hat{y}_t \ \hat{\Pi}_t \ \hat{R}_t \ \hat{\varepsilon}_t^S \ \hat{\varepsilon}_t^D \ \hat{y}_{t-1}]'$ and $\boldsymbol{\eta}_t$ constitutes the

vector of structural shocks $\boldsymbol{\eta}_t = [\eta_t^M \quad \eta_t^S \quad \eta_t^D]'$. The solution to this linear rational expectation model, derived on the basis of the algorithm proposed by Blanchard and Kahn (1980), can be expressed as the VAR model:

$$\mathbf{s}_t = \mathbf{B}_1(\boldsymbol{\Theta})\mathbf{s}_{t-1} + \mathbf{B}_2(\boldsymbol{\Theta})\boldsymbol{\eta}_t, \quad (30)$$

where \mathbf{s}_t denotes the vector of state variables $\mathbf{s}_t = [\hat{R}_t \quad \hat{\Pi}_t \quad \hat{y}_t \quad \hat{y}_{t-1} \quad \hat{\varepsilon}_t^S \quad \hat{\varepsilon}_t^D]'$.

For the purpose of estimating structural parameters of the model $\boldsymbol{\Theta}$ it is necessary to link the unobservable state variables with observable ones. In our application we relate state vector to three key macroeconomic variables of the U.S. economy: short term interest rates (R_t^{obs}), quarterly inflation (Π_t^{obs}) and quarterly growth rate of output (Y_t^{obs}). Consequently, a measurement equation is of the form:

$$\begin{bmatrix} R_t^{obs} \\ \Pi_t^{obs} \\ Y_t^{obs} \end{bmatrix} = \begin{bmatrix} \bar{R} \\ \bar{\Pi} \\ \bar{Y} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \end{bmatrix} \mathbf{s}_t, \quad (31)$$

where \bar{Y} is steady state growth rate of output, which is the sum of equilibrium growth rate of output per capita (g) and population growth rate.

The system consisting of the state equation (30) and the measurement equation (31) constitutes a state-space model and therefore we can apply the Kalman filtering to calculate the value of the likelihood function. The model includes twelve parameters incorporated in the vector $\boldsymbol{\Theta}$, three parameters that describe diagonal matrix of shock variance, ie. σ^M , σ^S and σ^D , and three parameters that pin down steady state values of output growth, inflation and nominal interest rates to the data, namely \bar{Y} , $\bar{\Pi}$ and \bar{R} . As in Ireland (2004) we relaxed model's assumption that $\bar{R} = g\bar{\Pi}/\beta$, which is not confirmed by the historical observations. Instead, we regard \bar{R} as the additional parameter to be estimated. Moreover, we decided to fix two of the structural parameters. The discount factor β was set at 0.995 and the inverse of Frisch elasticity φ was set at 2. The remaining coefficients were estimated by maximizing the value of the likelihood function of the state-space model.

2.2. Vector Autoregression model

VAR models, introduced by Sims in 1980, have been widely applied in macroeconomics, both for policy analysis and forecasting. Recently, they have additionally been used as a benchmark in evaluating performance of DSGE models. For that reason, we also investigate the quality of forecasts from a trivariate VAR model.

We analyse the vector of three U.S. macroeconomic variables that are exactly the same as those which are explained by the DSGE model introduced in the previous subsection, namely short term interest rates (R_t^{obs}), quarterly inflation (Π_t^{obs}) and quarterly growth rate of output (Y_t^{obs}). That means that for a vector $\mathbf{X}_t = [R_t^{obs} \quad \Pi_t^{obs} \quad Y_t^{obs}]'$ we estimate the following VAR model:

$$\mathbf{X}_t = \mathbf{\Gamma}_0 + \sum_{i=1}^P \mathbf{\Gamma}_i \mathbf{X}_{t-i} + \boldsymbol{\varepsilon}_t, \quad (32)$$

where $\mathbf{\Gamma}_0$ is a vector of intercepts, $\mathbf{\Gamma}_i$ are matrices of autoregressive coefficients, P is the lag order and $\boldsymbol{\varepsilon}_t$ is a vector of residuals. The residuals are assumed to follow a multivariate white noise processes, so that $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s') = \mathbf{0}$ for $t \neq s$ and $E(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \mathbf{\Omega}$ for $t = s$, where $\mathbf{\Omega}$ is a symmetric, positive defined variance-covariance matrix. The choice of the optimal lag order was based on the final prediction error criterion, where the maximum available lag was set as five.

2.3. The Survey of Professional Forecasters

The SPF is the oldest quarterly survey of macroeconomic forecasts in the United States. The survey, which was launched in 1968, was initially elaborated by the American Statistical Association and the National Bureau of Economic Research. Subsequently, in 1990 it was taken over by the Federal Reserve Bank of Philadelphia, which has conducted the SPF since then. The survey is carried out in regular three month intervals and concerns dozens of macroeconomic variables, among them output, inflation and short term interest rates. Since each of the thirty anonymous respondents to the survey is producing regular economic forecasts as part of his or her professional activity, the survey was baptized as the Survey of

Professional Forecasters. The results of the survey are published quarterly on the Philadelphia Fed web page.³

As discussed by Croushore (2006), the survey forms are sent at the end of the first month of each quarter, just after the advance release of the GDP data for the previous period, and are returned in the middle of the next month, i.e. before the data are revised. Nevertheless, the forecasters while formulating their predictions concerning the U.S. economy may use some additional information. Among others, they know about the leading indicators and business surveys for the previous month or about the current situation on the financial markets. Bearing that in mind, it seems obvious that the SPF has an advantage in forecasting output, prices and particularly interest rates in comparison to the above described estimated models, particularly in a short term horizon.

In our analysis we focus on the median forecasts of the three following variables: the quarterly GDP growth rate, the GDP price index inflation and the three month Treasury bill yield. All these variables are forecasted by the SPF up to four quarters ahead, where one-step forecast concerns the quarter when the survey is carried out. For instance, in the case of forecasts for the period 1994:01-1994:04 we evaluate the outcome of the survey from the first quarter of 1994. We proceed by using proper surveys from the period 1994:01-2005:03, which enables us to obtain time series of forecasts for the three analysed variables at four different forecasting horizons.

3. The data

We consider three quarterly sampled variables for the U.S. economy: the three month Treasury bill yield, quarterly growth rate of the seasonally adjusted real GDP and quarterly growth rate of the seasonally adjusted GDP price index. These variables represent models' short term interest rates, output growth and inflation, respectively. Since the goal of the analysis is to compare forecasting performance of the SPF to forecasts deriving from the two estimated models and due to the fact that time series are revised over time, the use of the recent available data would lead to favouring investigated models for the reasons discussed by Croushore (2006). The natural way to tackle this problem is to base the estimation on real-time data, which increases comparability of the forecasting errors as all types of predictions are formulated on the basis of the same data set.

³ See: <http://www.phil.frb.org/econ/spf/index.html>.

By the term “real-time data” we understand values of macroeconomic time series available to a researcher on a given date in the past. Following Croushore and Stark (2001a), we will refer henceforth to the date of observation as a “vintage,” and to the collection of time series from various vintages as a “real-time data set.” In our analysis, real-time data for the GDP and GDP price index growth rates are taken from the Philadelphia Fed Real-Time Data Set for Macroeconomists.⁴ Since the vintages are set as the middle day of the middle month of each quarter, each vintage includes the advance GDP data for the previous period. This means that these real-time data match up exactly the data that are available to the SPF. In case of the three month Treasury bill yield, the time series are not revised over time, and therefore the latest available data are the same as the real-time ones.⁵

The out-of-sample forecast performance is analysed for horizons ranging from one up to four quarters ahead and is evaluated with the use of the data from the period of 1994:01-2006:02. Prior to that, the DSGE and the VAR models are estimated on the basis of the most recent sixty quarterly observations for the vintage date, which is the period of forecast formulation. For instance, forecasts elaborated in the first quarter of 1994 for the period 1994:01-1994:04 are generated from the DSGE and the VAR models estimated from a sample covering the data span from 1979:01 to 1993:04 given by the vintage of the first quarter of 1994. The next pair of models is estimated with the use of the data from the vintage of the second quarter 1994 for the period of 1979:02-1994:01. These models are subsequently applied to forecast the U.S. economy up to four quarters ahead, i.e. for the period 1994:02-1995:01. This procedure is repeated for each quarter of the period of 1994:01-2005:03, which means that we calculate forty seven forecasts for each forecast horizon, each model and each of the analysed variables. These forecasts are then compared to the actual values in order to compute the forecast errors. Since the analysis is conducted in a real-time environment, a question arises which observations can be considered as the actual ones. We tackle this problem in two ways. Firstly, we analyse the forecast performance by taking into account the latest available data set, i.e. from the vintage of the third quarter of 2006, as the realization of variables. In the second case we compare the forecasts with real-time data available one year after a given date of the vintage used in estimation. We provide more detailed discussion of these issues in the next section.

⁴ The data are available on the Philadelphia Fed web page: <http://www.phil.frb.org/econ/forecast/reaindex.html>.

⁵ The data are available on the Fed web page: <http://www.federalreserve.gov/releases/>.

4. Results

In this section we present the results of the analysis aimed at comparing the out-of-sample forecast performance of the SPF, the VAR model and the DSGE model for the short term interest rate, the output growth and the inflation at horizons up to four quarters. Since the analysis is conducted in the real-time data environment, while calculating forecast errors we must decide on what to use as “actuals” for the forecasted variables. As mentioned in the previous sections, we evaluate the quality of forecasts in two variants. Firstly, we consider the latest available data set, i.e. the third quarter of 2006 vintage, as the realization of variables. Secondly, we compare the forecasts with real-time data available one year after the date of the vintage used in estimation. We label the former case as “latest available” and the latter one as “one year after estimation.”

We start out by examining whether the forecasts are biased. For the three analysed methods, three variables and four forecast horizons we regress the “actuals” (X_t) on the forecasts (X_t^F), namely we estimate the following models:

$$X_t = \alpha_0 + \alpha_1 X_t^F + \varepsilon_t. \quad (33)$$

Subsequently, we test the null hypothesis that the constant term is zero ($\alpha_0 = 0$) and the slope coefficient is unity ($\alpha_1 = 1$), which if accepted indicates that the forecast is unbiased. For that purpose we apply the Wald Chi-squared test corrected for heteroskedasticity and autocorrelation of the residuals. The adequate covariance matrix is estimated in line with the Newey and West (1987) procedure: we use the modified Bartlett kernel, where the truncation lag is dependent on the number of observations as proposed by Newey and West (1994).

The coefficient estimates with corresponding corrected standard errors for model (33), the coefficient of determination R^2 and the p -value for the unbiasedness hypothesis test are shown in Table 1 and Table 2, for the “latest available” data set and the “one year after estimation” data set cases, respectively. The results indicate that the short term interest rate forecasts from the VAR and DSGE models are unbiased, while in the case of the SPF one-quarter horizon forecast the null hypothesis is rejected at 1% significance level. The forecasts of the output growth and inflation turned out to be imprecise: the relevant coefficients of determination are low and hardly exceed ten percent. Moreover, inflation forecasts are biased in almost all cases. The only exception is the SPF one-quarter ahead forecast, if the “one year

after estimation” data set is used as “actuals.” Finally, in case of the output growth forecasts, the unbiasedness hypothesis cannot be rejected for the DSGE model at three- and four-quarter horizons, for the VAR model at three-quarter horizon and for the SPF at one-quarter horizon, if the “latest available” data set is considered.

We proceed our analysis by comparing the mean errors (ME), the mean absolute errors (MAE) and the root mean squared errors (RMSE) of forecasts. The corresponding measures of forecast performance are reported in Table 3. According to the results, the RMSEs of forecasts for the short term interest rate are the lowest for the SPF and the highest for the VAR model. The superiority of the SPF over the remaining two methods is evident especially for the one-quarter ahead forecast, which should not be surprising as the Professional Forecasters know about interest rate changes that occurred in the first half of the quarter for which the forecast is elaborated. In the case of the output growth forecasts the DSGE model is characterized by the lowest RMSEs at three- and four-quarter horizons, while the SPF outperforms the DSGE and VAR models if one-quarter ahead forecasts are considered. With regard to the inflation forecasts, we find that both in the “latest available” and “one year after estimation” data set cases the SPF forecasts are characterized by the lowest RMSEs among competing methods at all horizons. Moreover, the RMSEs of forecasts from the DSGE model turned out to be lower than those from the VAR model.

While the RMSE is widely used in evaluating forecast performance of a given method it does not allow to indicate if one method is statistically better than another. We cope with this issue by employing the Harvey-Leybourne-Newbold (1997) modification of the Diebold-Mariano (1995) test for the equal forecast accuracy from two competing models. Given time series $\{e_{1,t}\}_{t=1}^T$ and $\{e_{2,t}\}_{t=1}^T$ of the forecast errors at a given horizon h from two models, Diebold and Mariano define a loss differential as $d_t = e_{1,t}^2 - e_{2,t}^2$ for $t = 1, 2, \dots, T$ and test the null hypothesis that it is equal to zero, namely $E(d_t) = 0$. If this hypothesis is rejected then the model characterized by smaller mean of squared forecast errors is significantly superior to the other one. The proposed statistic takes the following form:

$$DM = \frac{\bar{d}}{\sqrt{\frac{2\pi \hat{S}_d(0)}{T}}}, \quad (34)$$

where $\bar{d} = T^{-1} \sum_{t=1}^T d_t$ is the sample mean of the loss differential. The expression $\hat{S}_d(0) = \hat{\gamma}_0^d + 2 \sum_{k=1}^{h-1} \hat{\gamma}_k^d$ constitutes nonparametric consistent estimate of the spectrum of the loss differential at frequency zero, where the k -th sample autocovariance of d_t is given by $\hat{\gamma}_k^d = T^{-1} \sum_{t=k+1}^T (d_t - \bar{d})(d_{t-k} - \bar{d})$. Under the null hypothesis the DM statistic has an asymptotic standard normal distribution. In case of small samples, Harvey, Leybourne and Newbold (1997) proposed an adjusted DM statistic, which for a given forecast horizon h is equal to:

$$HLN = \sqrt{\frac{T+1-2h+h(h-1)/T}{T}} DM \quad (35)$$

and has the t -Student distribution with $T-1$ degrees of freedom.

Up to this point we have indicated that the RMSEs of forecasts from the SPF are lower than those from the VAR and DSGE models if forecasts of the short term interest rate are considered. This conclusion is partly confirmed by the Harvey-Leybourne-Newbold test, reported in Table 4, which shows that at 5% significance level the SPF forecasts are superior to the forecasts from the VAR model at horizons up to three quarters, and to the forecasts from the DSGE model at one- and two-quarter horizons. At three- and four-quarter horizons the null hypothesis cannot be rejected, indicating equal forecast accuracy of the SPF and the DSGE model. Moreover, we cannot reject the null hypothesis that the quality of the short term interest rate forecasts from the DSGE and VAR models are different. The comparison of the RMSEs of the output growth forecasts shows that at two-, three- and four-quarter horizons the forecast accuracy of the three analysed methods is not significantly different. This means that it would be unwarranted to claim that the DSGE model outperforms the SPF in forecasting the output growth. In the case of the one-quarter ahead forecasts, if the “latest available” data set is taken as “actuals,” we reject the null which means that the DSGE and the SPF outperform the VAR model. As regards inflation forecasts for all horizons, at the 5% significance level we could not reject the null that forecasts from the SPF and the DSGE model are different from each other, both for the “latest available” and “one year after estimation” data set cases. The results indicate, however, that the SPF outperforms the VAR model at one- and two-quarter horizons. Finally, the comparison of the DSGE and the VAR model forecast accuracy yields no significant difference: the null of equal forecast accuracy cannot be rejected at any reasonable significance level for all considered forecast horizons.

5. Conclusions

In the paper we have compared the quality of forecasts from the DSGE and VAR models as well as from the SPF in the case of three key macroeconomic variables for the U.S. economy. First of all, we tested which methods generate unbiased forecasts. Subsequently, we analysed the forecast accuracy of the presented methods by comparing various *ex post* forecast error measures like MA, MAE and RMSE. Finally, we applied the Harvey-Leybourne-Newbold (1997) test to check if one method can significantly outperform the other one in forecasting the U.S. economy. Since we carried out the analysis in the real-time context, we controlled comparability of the information available to the SPF with the time series used to estimate coefficients of the DSGE and VAR models. According to our best knowledge this is the first study that compares the forecast errors from a DSGE model with those from the SPF in a real-time environment.

The obtained results show that the short term interest rate forecasts were unbiased in case of all methods and horizons, except from the one-quarter ahead forecast from the SPF. In contrast, all inflation forecasts turned out to be biased. Moreover, these forecasts come out to be imprecise: the relevant coefficients of determination were always below ten percent. In the case of the output growth forecasts, the only unbiased predictions were generated by the DSGE model at three- and four-quarter horizons, and the VAR model at three-quarter horizon. Comparison of the forecasts RMSEs led to the result that the DSGE model outperforms the SPF in three- and four-quarter ahead forecasts of the output growth. In all other cases we found that the SPF is superior to both the DSGE and the VAR models. The HLN test of the null hypothesis about equal forecast accuracy showed that in some cases forecasts of the inflation and the short term interest rate from the SPF are significantly better than those from the DSGE and the VAR models. However, when comparing the output growth forecasts from the SPF and from the DSGE model the null could not be rejected at any forecast horizon, indicating that the differences in the corresponding RMSEs are not statistically significant. The general picture that emerges from the above analysis is that the proposed DSGE model is not able to significantly outperform the SPF in forecasting output growth, inflation and interest rates in the United States. We found, however, that the DSGE model generates forecasts which are very close in accuracy to the SPF predictions. Moreover, the DSGE model occurred to perform better than the trivariate VAR model.

Overall, the results of the analysis are mixed. We have shown that the small scale DSGE model can produce forecasts whose accuracy is in some cases comparable to the

forecasts from the SPF, and in some cases superior to the forecasts from the VAR model. Clearly, additional research is required to document the out-of-sample performance of DSGE models. First of all, the structure of the DSGE model presented in this article is relatively simple and hence forecasting proprieties of more complex DSGE model could be studied. Secondly, since the quality of forecasts from DSGE models depends on the choice of the estimation technique, other estimators such as bayesian or generalized method of moments ones could be used. Finally, forecast accuracy of DSGE models could be compared to a larger group of methods than the SPF and VAR models. Nonetheless, we hope that our findings constitute an argument in favour of increased use of DSGE models in macroeconomic forecasting.

References

- Blanchard, O., Kahn, C. (1980), "The Solution of Linear Difference Models Under Rational Expectations," *Econometrica*, Vol. 48, p. 1305-1311.
- Calvo, G. (1983), "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, Vol. 12, p. 383-398.
- Christiano, L., Eichenbaum, M., Evans, C. (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, Vol. 113, Iss. 1, p. 1-45.
- Clark, T., McCracken, M. (2006), "Forecasting with Small Macroeconomic VARs in the Presence of Instabilities," Working Paper No. 06-09, Federal Reserve Bank of Kansas City.
- Croushore, D. (2006), "An Evaluation of Inflation Forecasts From Surveys Using Real-Time Data," Working Paper No. 06-19, Federal Reserve Bank of Philadelphia.
- Croushore, D., Stark, T. (2001a), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics*, Vol. 105, p. 111-130.
- Croushore, D., Stark, T. (2001b), "Forecasting with a Real-Time Data Set for Macroeconomists," Working Paper No. 01-10, Federal Reserve Bank of Philadelphia.
- Del Negro, M., Schorfheide, F., Smets, F., Wouters, R. (2005), "On the Fit and Forecasting Performance of New Keynesian Models," CEPR Discussion Paper Series No. 4848, Centre for Economic Policy Research, London.
- Diebold, F. X., Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, Vol. 13, No. 3, p. 253-263.
- Dixit, A., Stiglitz, J. (1977), "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, Vol. 67, p. 297-308.
- Edge, R., Kiley, M., Laforte, J. (2006), "A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model," http://www.eabcn.org/workshops/zurich_2006/documents/Laforte.pdf.

- Harvey, D., Leybourne, S., Newbold, P. (1997), "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, Vol. 13, No. 2, p. 281-291.
- Ireland, P. (2004), "Technology Shocks in the New Keynesian Model," NBER Working Paper No. 10309, National Bureau of Economic Research, Cambridge.
- Kydland, F., Prescott, C. (1982), "Time To Build and Aggregate Fluctuations," *Econometrica*, Vol. 50, p. 1345-1370.
- Newey, W. K., West, K. D. (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *The Review of Economic Studies*, Vol. 61, No. 4, p. 631-653.
- Newey, W. K., West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, Vol. 55, No. 3, p. 703-708.
- Obstfeld, M., Rogoff, K. (1995), "Exchange Rate Dynamics Redux," *The Journal of Political Economy*, Vol. 103, No. 3, p. 624-660.
- Orphanides, A., Williams, J. (2002), "Robust monetary policy rules with unknown natural rates," *Brookings Papers on Economic Activity*, Vol. 2, p. 63-118.
- Rudebusch, G. (2002), "Term structure evidence on interest-rate smoothing and monetary policy inertia," *Journal of Monetary Economics*, Vol. 49, p. 1116-1186.
- Sims, Ch. (1980), "Macroeconomics and Reality," *Econometrica*, Vol. 48, p. 1-48.
- Smets, F., Wouters, R. (2004), "Forecasting With a Bayesian DSGE Model: An Application to the Euro Area," Working Paper No. 389, European Central Bank, Frankfurt.
- Smets, F., Wouters, R. (2003), "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, Vol. 1, Iss. 5, p. 1123-1175.
- Taylor, J. (1993), "Discretion versus policy rules in practice," *Carnegie-Rochester Series on Public Policy*, Vol. 39, p. 195-214.

Appendix

Table 1. Tests of the hypothesis about forecast unbiasedness – “latest available” data set case

h	<i>Short term interest rate</i>				<i>Output growth</i>				<i>Inflation</i>			
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	p -value	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	p -value	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	p -value
SPF												
1	0.002 (0.030)	0.987 (0.007)	0.997	0.004	1.206 (0.635)	0.774 (0.189)	0.189	0.103	1.488 (0.518)	0.284 (0.216)	0.033	0.002
2	0.007 (0.131)	0.964 (0.031)	0.946	0.174	2.924 (1.026)	0.103 (0.301)	0.001	0.011	2.196 (0.582)	-0.057 (0.247)	0.001	0.000
3	0.019 (0.311)	0.937 (0.071)	0.830	0.305	4.943 (1.107)	-0.608 (0.413)	0.037	0.000	2.612 (0.676)	-0.229 (0.282)	0.013	0.000
4	0.036 (0.566)	0.909 (0.127)	0.656	0.337	4.882 (1.190)	-0.577 (0.405)	0.032	0.000	3.379 (0.713)	-0.560 (0.290)	0.090	0.000
VAR												
1	0.232 (0.243)	0.969 (0.055)	0.912	0.497	3.780 (1.002)	-0.146 (0.228)	0.005	0.000	2.008 (0.404)	0.030 (0.177)	0.001	0.000
2	0.735 (0.574)	0.845 (0.128)	0.705	0.440	2.605 (1.112)	0.178 (0.281)	0.009	0.006	2.119 (0.402)	-0.022 (0.200)	0.000	0.000
3	0.965 (0.692)	0.761 (0.161)	0.555	0.331	1.100 (1.747)	0.547 (0.448)	0.047	0.150	2.203 (0.388)	-0.049 (0.194)	0.002	0.000
4	1.304 (0.897)	0.654 (0.211)	0.377	0.257	1.761 (1.610)	0.376 (0.373)	0.024	0.009	2.644 (0.382)	-0.270 (0.179)	0.065	0.000
DSGE												
1	-0.057 (0.158)	1.020 (0.040)	0.942	0.874	2.168 (1.193)	0.319 (0.294)	0.016	0.023	1.655 (0.407)	0.206 (0.161)	0.029	0.000
2	0.065 (0.366)	0.981 (0.091)	0.820	0.978	3.323 (1.465)	-0.032 (0.414)	0.000	0.038	2.164 (0.509)	-0.040 (0.200)	0.001	0.000
3	0.348 (0.597)	0.894 (0.145)	0.651	0.758	1.736 (1.362)	0.454 (0.445)	0.016	0.444	2.245 (0.510)	-0.059 (0.194)	0.002	0.000
4	0.768 (0.826)	0.773 (0.195)	0.471	0.478	3.617 (2.138)	-0.124 (0.723)	0.001	0.177	3.104 (0.562)	-0.414 (0.211)	0.079	0.000

Notes: Table presents the coefficient estimates of the regression (33). In order to correct for heteroskedasticity and autocorrelation we applied the Newey-West procedure using modified Bartlett kernel with truncation lag fixed at three. The values in parenthesis denote corrected standard errors. Reported p -values relate to the test of the null hypothesis that the forecast is unbiased.

Table 2. Tests of the hypothesis about forecast unbiasedness – “one year after estimation” data set case

<i>h</i>	<i>Short term interest rate</i>				<i>Output growth</i>				<i>Inflation</i>			
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	<i>p-value</i>	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	<i>p-value</i>	$\hat{\alpha}_0$	$\hat{\alpha}_1$	R^2	<i>p-value</i>
SPF												
1	0.002 (0.030)	0.987 (0.007)	0.997	0.004	1.591 (0.658)	0.666 (0.228)	0.161	0.007	0.177 (0.468)	0.870 (0.204)	0.256	0.517
2	0.007 (0.131)	0.964 (0.031)	0.946	0.174	2.614 (0.865)	0.268 (0.279)	0.011	0.009	1.307 (0.539)	0.294 (0.231)	0.025	0.002
3	0.019 (0.311)	0.937 (0.071)	0.830	0.305	3.855 (0.939)	-0.165 (0.358)	0.003	0.000	1.402 (0.594)	0.244 (0.244)	0.014	0.001
4	0.036 (0.566)	0.909 (0.127)	0.656	0.337	4.244 (1.161)	-0.366 (0.397)	0.017	0.001	2.272 (0.701)	-0.180 (0.277)	0.009	0.000
VAR												
1	0.232 (0.243)	0.969 (0.055)	0.912	0.497	3.344 (0.742)	0.005 (0.189)	0.000	0.000	1.335 (0.278)	0.327 (0.121)	0.071	0.000
2	0.735 (0.574)	0.845 (0.128)	0.705	0.440	2.387 (1.122)	0.287 (0.270)	0.027	0.008	1.683 (0.343)	0.127 (0.164)	0.012	0.000
3	0.965 (0.692)	0.761 (0.161)	0.555	0.331	1.341 (1.239)	0.528 (0.312)	0.050	0.126	1.641 (0.286)	0.157 (0.149)	0.020	0.000
4	1.304 (0.897)	0.654 (0.211)	0.377	0.257	1.780 (1.436)	0.363 (0.362)	0.031	0.010	1.950 (0.262)	-0.042 (0.133)	0.001	0.000
DSGE												
1	-0.057 (0.158)	1.020 (0.040)	0.942	0.874	2.284 (1.077)	0.314 (0.305)	0.018	0.079	1.044 (0.324)	0.449 (0.131)	0.116	0.000
2	0.065 (0.366)	0.981 (0.091)	0.820	0.978	4.152 (1.361)	-0.232 (0.375)	0.006	0.004	1.486 (0.442)	0.202 (0.174)	0.019	0.000
3	0.348 (0.597)	0.894 (0.145)	0.651	0.758	2.431 (1.446)	0.292 (0.445)	0.008	0.239	1.726 (0.432)	0.093 (0.172)	0.004	0.000
4	0.768 (0.826)	0.773 (0.195)	0.471	0.478	4.036 (2.056)	-0.266 (0.692)	0.005	0.093	2.273 (0.426)	-0.171 (0.156)	0.012	0.000

Notes: Table presents the coefficient estimates of the regression (33). In order to correct for heteroskedasticity and autocorrelation we applied the Newey-West procedure using modified Bartlett kernel with truncation lag fixed at three. The values in parenthesis denote corrected standard errors. Reported *p*-values relate to the test of the null hypothesis that the forecast is unbiased.

Table 3. Out-of-sample forecast evaluation

h	<i>Short term interest rate</i>			<i>Output growth</i>			<i>Inflation</i>		
	<i>SPF</i>	<i>VAR</i>	<i>DSGE</i>	<i>SPF</i>	<i>VAR</i>	<i>DSGE</i>	<i>SPF</i>	<i>VAR</i>	<i>DSGE</i>
ME – “latest available” data set case									
1	-0.046	0.119	0.020	0.605	-0.269	-0.168	0.042	0.229	0.079
2	-0.136	0.172	-0.008	0.392	-0.205	-0.157	0.000	0.217	-0.063
3	-0.234	0.071	-0.062	0.387	-0.657	-0.051	-0.077	0.238	-0.150
4	-0.343	-0.025	-0.130	0.350	-0.666	0.046	-0.106	0.222	-0.225
ME – “one year after estimation” data set case									
1	-0.046	0.119	0.020	0.703	-0.172	-0.070	-0.086	0.101	-0.049
2	-0.136	0.172	-0.008	0.548	-0.049	-0.002	-0.159	0.059	-0.222
3	-0.234	0.071	-0.062	0.554	-0.490	0.117	-0.253	0.062	-0.326
4	-0.343	-0.025	-0.130	0.317	-0.699	0.013	-0.364	-0.036	-0.483
MAE – “latest available” data set case									
1	0.075	0.355	0.295	1.499	1.937	1.702	0.640	0.828	0.747
2	0.285	0.671	0.530	1.731	1.706	1.685	0.738	0.844	0.797
3	0.563	0.879	0.756	1.767	1.662	1.538	0.773	0.923	0.848
4	0.820	1.062	0.964	1.780	1.768	1.620	0.881	1.023	0.968
MAE – “one year after estimation” data set case									
1	0.075	0.355	0.295	1.472	1.742	1.634	0.561	0.732	0.682
2	0.285	0.671	0.530	1.574	1.639	1.637	0.677	0.806	0.762
3	0.563	0.879	0.756	1.667	1.522	1.533	0.723	0.837	0.782
4	0.820	1.062	0.964	1.579	1.545	1.513	0.861	0.926	0.949
RMSE – “latest available” data set case									
1	<u>0.099</u>	0.516	0.408	<u>1.910</u>	2.311	2.054	<u>0.820</u>	1.025	0.899
2	<u>0.419</u>	0.964	0.715	2.140	2.192	<u>2.116</u>	<u>0.896</u>	1.066	0.966
3	<u>0.740</u>	1.193	1.009	2.250	2.096	<u>2.013</u>	<u>0.922</u>	1.120	0.983
4	<u>1.056</u>	1.439	1.280	2.226	2.151	<u>2.072</u>	<u>1.027</u>	1.259	1.115
RMSE – “one year after estimation” data set case									
1	<u>0.099</u>	0.516	0.408	<u>1.878</u>	2.116	1.920	<u>0.722</u>	0.923	0.853
2	<u>0.419</u>	0.964	0.715	2.032	<u>2.021</u>	2.051	<u>0.867</u>	1.003	0.940
3	<u>0.740</u>	1.193	1.009	2.086	1.927	<u>1.911</u>	<u>0.880</u>	0.998	0.986
4	<u>1.056</u>	1.439	1.280	1.932	1.910	<u>1.820</u>	<u>1.042</u>	1.151	1.155

Notes: Underlined values are the minimum RMSEs for each variable and forecast horizon h .

Table 4. Equal forecast accuracy tests

<i>h</i>	<i>Short term interest rate</i>		<i>Output growth</i>		<i>Inflation</i>	
	<i>HLN statistic</i>	<i>p-value</i>	<i>HLN statistic</i>	<i>p-value</i>	<i>HLN statistic</i>	<i>p-value</i>
SPF vs. VAR – “latest available” data set case						
1	-3.130	0.003	-2.029	0.048	-2.968	0.005
2	-2.009	0.050	-0.248	0.805	-1.863	0.069
3	-2.031	0.048	0.555	0.581	-1.821	0.075
4	-1.779	0.082	0.329	0.743	-1.688	0.098
SPF vs. VAR – “one year after estimation” data set case						
1	-3.130	0.003	-1.073	0.289	-3.581	0.001
2	-2.009	0.050	0.053	0.958	-2.218	0.032
3	-2.031	0.048	0.618	0.539	-1.987	0.053
4	-1.779	0.082	0.076	0.940	-1.409	0.166
SPF vs. DSGE – “latest available” data set case						
1	-3.253	0.002	-0.852	0.399	-1.635	0.109
2	-2.275	0.028	0.174	0.862	-1.839	0.072
3	-1.627	0.111	1.372	0.177	-1.496	0.142
4	-1.389	0.172	0.894	0.376	-1.654	0.105
SPF vs. DSGE – “one year after estimation” data set case						
1	-3.253	0.002	-0.197	0.844	-1.843	0.072
2	-2.275	0.028	-0.126	0.900	-1.657	0.104
3	-1.627	0.111	1.032	0.307	-1.710	0.094
4	-1.389	0.172	0.641	0.525	-1.544	0.129
DSGE vs. VAR – “latest available” data set case						
1	-1.303	0.199	-2.124	0.039	-1.944	0.058
2	-1.096	0.279	-0.436	0.665	-1.294	0.202
3	-0.861	0.394	-0.457	0.650	-1.263	0.213
4	-0.755	0.454	-0.750	0.457	-1.010	0.318
DSGE vs. VAR – “one year after estimation” data set case						
1	-1.303	0.199	-1.971	0.055	-1.256	0.216
2	-1.096	0.279	0.186	0.854	-0.998	0.323
3	-0.861	0.394	-0.093	0.927	-0.126	0.901
4	-0.755	0.454	-0.505	0.616	0.031	0.976

Notes: Under the null hypothesis of equal forecast accuracy the HLN statistic has a *t*-Student distribution.