Evaluating point and density forecasts of DSGE models

Maik H. Wolters *
University of Kiel and Kiel Institute for the World Economy

August 10, 2013

Abstract

This paper investigates the accuracy of forecasts from four DSGE models for inflation, output growth and the federal funds rate using a real-time dataset synchronized with the Fed's Greenbook projections. Conditioning the model forecasts on the Greenbook nowcasts leads to forecasts that are as accurate as the Greenbook projections for output growth and the federal funds rate. Only for inflation the model forecasts are dominated by the Greenbook projections. A comparison with forecasts from Bayesian VARs shows that the economic structure of the DSGE models which is useful for the interpretation of forecasts does not lower the accuracy of forecasts. Combining forecasts of several DSGE models increases precision in comparison to individual model forecasts. Comparing density forecasts with the actual distribution of observations shows that DSGE models overestimate uncertainty around point forecasts.

Keywords: DSGE models, Bayesian VAR, forecasting, model uncertainty, forecast combination, density forecasts, real-time data, Greenbook

JEL-Codes: C53, E31, E32, E37

*Mailing address: Kiel Institute for the World Economy, Hindenburgufer 66, 24105 Kiel, Germany; maik.wolters@ifw-kiel.de. An earlier version of this paper circulated under the title "Forecasting under model uncertainty". I am grateful for helpful comments by Tim Oliver Berg, Tobias Cwik, Ester Faia, Stefan Gerlach, Lucrezia Reichlin, Sebastian Schmidt, Frank Schorfheide, Tomi Wang, Mike Wickens, Volker Wieland as well as seminar participants at Goethe University Frankfurt, Stanford University, the Federal Reserve Bank of Philadelphia, the Federal Reserve Board, the University of Kiel, the 2010 DIW macroeconometric workshop in Berlin, the 2010 IWH macroeconomic workshop in Halle, the 2010 ZEW macroeconomics conference in Mannheim, the 2010 RGS conference in Bochum, the 2011 SNDE conference in Washington DC, the 2011 Midwest Macro Meeting in Nashville, the 2011 CEF conference in San Francisco, the 2011 Dynare conference in Atlanta and the 2011 annual meeting of the Verein für Socialpolitik in Frankfurt.

1 Introduction

For a long time business cycle models with microeconomic foundations have been calibrated and used for policy simulations while atheoretical time series methods have been used to forecast macroeconomic variables. In recent years, several researchers have shown that estimated dynamic stochastic general equilibrium (DSGE) models can generate forecasts with an accuracy comparable to the one of nonstructural time series models (Smets and Wouters, 2004; Adolfson et al., 2007a; Smets and Wouters, 2007; Wang, 2009; Edge et al., 2010; Christoffel et al., 2011).[1] The advantage of using structural models is that an economically meaningful interpretation of the forecasts can be given.[2] All these studies analyse only one DSGE model at a time. Although all major central banks have included one or several DSGE models into their forecasting toolkit[3] a thorough assessment of forecasts from several structural models including a comparison to forecasts from sophisticated nonstructural time series methods and to professional forecasts has not been undertaken yet. Recent comparison studies of state of the art forecasting methods have been restricted to nonstructural econometric methods (c.f. Stock and Watson, 2002; Bernanke and Boivin, 2003; Forni et al., 2003; Marcellino et al., 2003; Faust and Wright, 2009; Hsiao and Wan, 2010).

In this paper, I carry out a detailed assessment of the forecasting accuracy of a suite of structural models. I consider models that cover to some extent the range of closed-economy DSGE models that have been used in academia and at policy institutions prior to the 2008/2009 financial crisis. The first model is a purely forward looking small-scale New Keynesian model with sticky prices that is analysed in detail in Woodford (2003). The second model by Fuhrer (1997) has a mainly backward looking demand side, while the Phillips curve is derived from expectation-based overlapping wage contracts. The third model is the medium-scale New Keynesian model by Smets and Wouters (2007). The fourth model is a version of the Fed's DSGE model by Edge et al. (2007) that features two production sectors with different technology growth rates and a demand side that is disaggregated into different consumption and investment categories. The parameters of the models are reestimated on three to eleven time series. Given this estimate, I compute a nowcast and forecasts up to five quarters into the future.

I use the same sample and real-time dataset as Faust and Wright (2009) who assess the forecasting accuracy of eleven nonstructural models. Therefore, my results are directly comparable to the forecasts from these models. The dataset includes data vintages for 123 FOMC meetings between 1984

[1]Edge and Gürkaynak (2010) study the absolute rather than relative forecasting accuracy of the DSGE model by Smets and Wouters (2007). They find that the forecasts contain little information about actual future macroeconomic dynamics, although statistical and judgmental forecasts perform equally poorly.

[2]Wieland and Wolters (2012) show examples how to structurally interpret forecasts from DSGE models.

[3]Examples include the FRB/EDO-model at the Fed (Edge et al., 2007, 2008, 2010), the New Area-Wide Model and the CMR model at the ECB (Christoffel et al., 2008; Smets et al., 2010), the ToTEM model at the Bank of Canada (Murchison and Rennison, 2006) or the Ramses model at the Riksbank (Adolfson et al., 2007b).

and 2000 and is perfectly synchronized with the Greenbook. Hence, the results can also be compared to a best practice benchmark given by the Greenbook projections.

The evaluation results of the point forecasts confirm the reasonable forecasting accuracy of DSGE models found in the above mentioned studies. Output growth has little persistence and is thus difficult to forecast in general. The DSGE models, the Greenbook projections and nonstructural forecasts perform as well as a simple autoregressive forecast, but not better. For the federal funds rate the Greenbook path is more accurate than the model forecasts for the first few quarters, but not for medium term horizons. The Greenbook inflation forecasts are more accurate than all DSGE model forecasts. A possible explanation is that the Fed considers the information content of more variables than processed by the DSGE models when forecasting inflation. To check this, I estimate a large Bayesian VAR as in Bańbura et al. (2010) that uses all variables from the dataset. Indeed, this model achieves medium term inflation forecasts that approach the precision of the Greenbook projections. Overall, the DSGE models work relatively well for medium term horizon forecasts and forecasting accuracy tends to increase with the number of time series used for the estimation of the models.

Good forecasts are in general based not only on good forecasting methods, but also on an accurate assessment of the current state of the economy. The Fed's great efforts to evaluate the current state of the economy are reflected in the accuracy of the Greenbook nowcasts. Sims (2002) suggests that this accurate data basis is a main reason for the precise Greenbook projections. Indeed, appending the Greenbook nowcast to the data and using it as a starting point for DSGE model forecasts increases forecasting precision a lot for persistent variables like inflation and the federal funds rate.

To isolate the role played by information and the role played by the discipline of economic theory in determining forecasting accuracy I estimate several Bayesian VARs that contain the same variables as the DSGE models. Forecasts of DSGE models and Bayesian VARs of the same size show a similar precisions so that the economic structure of the DSGE models that is helpful in interpreting the forecast is no hindrance in achieving accurate forecasts.

To assess the importance of data revisions in forecast evaluation, I repeat the forecasting exercise using revised data. The ranking of the different forecasting models does not change much when using revised instead of real-time data.

Another robustness check consists of dividing the evaluation period into subsamples. Here I find that no DSGE model continuously dominates the other models. The problem of instability in the performance and predictive content of different forecasting methods is well known and surveyed in Rossi (2012). Timmermann (2006) provides an overview about model averaging methods and finds that weighted forecasts from several nonstructural models outperform forecasts from individual models. Combining several models provides a hedge against model uncertainty when it is not possible to identify a single model that consistently dominates the forecasting accuracy of other models. In the context of DSGE models several researchers have combined different forecasting approaches to achieve higher forecasting accuracy. Ingram and Whiteman (1994) use DSGE models to construct priors for a VAR. Del Negro and Schorfheide (2004) extend that approach to DSGE-VARs. Gerard and Nimark (2008) and Bache et al. (2011) combine forecasts from VARs and a DSGE model, Waggoner and Zha (2010) combine a DSGE model with a Bayesian VAR and Geweke and Amisano (2012) include in addition a dynamic factor model (DFM) into the forecasting pool. I evaluate the forecasting accuracy of combined forecasts from the four DSGE models and consider several simple and sophisticated model averaging schemes to compute weighted forecasts. I find that averaged forecasts that give considerable weight to several models have a higher accuracy than forecasts from individual models and lead for medium horizons to forecasts comparable to the Greenbook projections. A simple average of the forecasts of all four DSGE models is in many cases the most accurate combination method.

While point forecasts are interesting, economists are concerned about the uncertainty surrounding these. Answers to questions like what is the probability of output growth being above 2% while inflation is between 0 and 2% are of interest to policy makers and DSGE models estimated with Bayesian techniques are suitable to answer these questions. However, for those answers to be meaningful, density forecasts of DSGE models should be a realistic description of actual uncertainty. So far the literature on the evaluation of DSGE model forecasts has focused on the evaluation of point forecasts with Herbst and Schorfheide (2012) being the only exception.[4] I derive density forecasts for the DSGE models that take into account parameter uncertainty and uncertainty about future economic shock realizations. I find that all model forecasts overestimate actual uncertainty, i.e. density forecasts are very wide when compared with the actual distribution of data. A reason might be the tight restrictions imposed on the data. If the data rejects these restrictions, large shocks are needed to fit the models to the data resulting in high shock uncertainty (see also Gerard and Nimark, 2008). In addition, estimating the models on the highly volatile data of the early 1980s and evaluating the forecasts on the Great Moderation sample puts the models to a very hard test and the changes in volatility contribute to the overestimation of uncertainty.

The remainder of this paper proceeds as follows. Section 2 outlines the different macroeconomic models that are used to compute forecasts. Section 3 gives an overview of the dataset. Section 4 describes the estimation and forecasting methodology. Section 5 evaluates point forecasts from the individual models and compares them to Greenbook projections and nonstructural forecasts. Section 6 provides a comparison of the accuracy of weighted forecasts. Section 7 evaluates density forecasts of the four DSGE models. Section 8 summarizes the findings and concludes.

[4]Their paper has been written simultaneously with my paper.

2 Forecasting models

I consider four different DSGE models of the U.S. economy. The models are chosen to broadly reflect the variety of DSGE models used in academia and at policy institutions prior to the 2007/2008 financial crisis.[5]

The first model is a version of the standard small New Keynesian model as used in Del Negro and Schorfheide (2004). The New Keynesian model is described, e.g., in Goodfriend and King (1997) and Rotemberg and Woodford (1997). It is often referenced to be the workhorse model in modern monetary economics and a comprehensive analysis is presented in the monograph of Woodford (2003). The model consists of three main equations: an IS curve, a monetary policy rule and a Phillips curve. The expectational IS curve can be derived from the behavior of optimizing and forward looking representative households that have rational expectations. Together with a monetary policy rule it determines aggregate demand. The New Keynesian Phillips curve determines aggregate supply and can be derived from monopolistic firms that face sticky prices. Wang (2009) shows that the small number of frictions is sufficient to provide reasonable output growth and inflation forecasts.

The second model is also a small-scale model of the U.S. economy described in Fuhrer (1997). It differs from the standard New Keynesian model with respect to the degree of forward lookingness and the specification of sticky prices. Aggregate demand is determined by a reduced form backward looking IS curve together with a monetary policy rule. Aggregate supply is modelled via overlapping wage contracts: agents care about real wage contracts relative to those negotiated in the recent past and those that are expected to be negotiated in the near future (see Fuhrer and Moore, 1995a,b). The aggregate price level is a constant mark-up over the aggregate wage rate. The resulting Phillips curve depends on current and past demand and expectations about future demand.

The third model is a medium-scale DSGE model by Smets and Wouters (2007). They extend the small New Keynesian model to explain more data series. It is a closed economy model that incorporates physical capital in the production function and capital formation is endogenized. Labor supply is modelled explicitly. Nominal frictions include sticky prices and wages as well as inflation and wage indexation. Real frictions include consumption habit formation, investment adjustment costs and variable capital utilization. The model includes equations for consumption, investment, price and wage setting as well as several identities.

The fourth model is a medium-scale DSGE model that has been developed at the Federal Reserve Board by Edge et al. (2008) and is called the FRB/EDO (Estimated Dynamic Optimized) model. The model features two production sectors, which differ with respect to the pace of technological

[5]To explain the recent financial crisis it is necessary to add a financial sector to the models (see e.g. Del Negro et al., 2013, for an extended model that can capture the large GDP drop during the Great Recession). I use models without a financial sector as my sample ends in 2000.

progress. This structure can capture the different growth rates and relative prices observed in the data. Accordingly, the expenditure side is disaggregated as well. It is divided into business investment and three categories of household expenditure: consumption of non-durables and services, investment in durable goods and residential investment. The model is able to capture different cyclical properties in these four expenditure categories. The model is documented in detail in Edge et al. (2007).

Table 1 summarizes the most important features of the four structural models. The models are abbreviated in the following: Del Negro & Schorfheide (DS), Fuhrer & Moore (FM), Smets & Wouters (SW), FRB/EDO model by Edge, Kiley & Laforte (EDO). The DS and FM models are estimated on the three key variables output growth, inflation and the federal funds rate. The SW model is estimated on seven variables: the three key variables and wages, hours worked, consumption and investment. The EDO model is estimated on eleven time series: output growth, inflation, the federal funds rate, consumption of non-durables and services, consumption of durables, residential investment, business investment, hours, wages, inflation for consumer nondurables and services and inflation for consumer durables. The method of estimating the structural parameters also varies across the models: I adapt the methodology used by the original authors and use maximum likelihood estimation for the FM model while Bayesian estimation is used to estimate the other models. For the priors, I use the ones in the original research referenced in table 1.

Table 1: Model overview

Model	Type	Observable Variables	Reference
DS	Small-scale microfounded forward looking New Keynesian Model	3: output growth, inflation, interest rate	Del Negro and Schorfheide (2004)
FM	Small-scale model with overlapping real wage contracts and a backward looking IS curve	3: output growth, inflation, interest rate	Fuhrer (1997)
SW	Medium-scale DSGE-model with many nominal and real frictions as used by policy institutions	7: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate	Smets and Wouters (2007)
EDO	Large-scale DSGE-model developed at the Federal Reserve Board. Two production sectors with different technology growth rates. The demand side is disaggregated into four categories	11: output growth, inflation, interest rate, consumption of non-durables and services, consumption of durables, residential investment, business investment, hours, wages, inflation for consumer non-durables and services, inflation for consumer durables	Edge et al. (2008)

Notes: Type: short classification of the models according to the main modeling assumptions; Observable Variables: the number and names of the observable variables; Reference: original reference that is closest to the implemented version in this paper.

3 A real-time dataset

I use the real-time dataset described in Faust and Wright (2009). These historical data vintages have been used by the Federal Reserve staff to compute the Greenbook projections. Thus, the dataset is perfectly synchronized with the Greenbook. The number of variables varies over the different vintages and the number of series in each vintage ranges from 47 to 80, with an average of 67. The dataset contains data vintages for 145 FOMC meetings from March 1980 to December 2000, while the different data series start in 1960.

While some of the nonstructural forecasting models considered in Faust and Wright (2009) and the large Bayesian VAR used here for comparison can process as many data series as available, the structural models considered in this paper use only a small subset of the available time series varying from three to eleven variables to estimate the different models. Still some variables for the EDO model are not available in the dataset. Therefore, I have added the necessary real-time data series from the Federal Reserve Bank of St. Louis' Alfred database. To each data vintage I add only observations that would have been available at the Greenbook publication date.

There is a trade-off between using a long sample to get precise parameter estimates and leaving out a fraction of past data to get rid off structural breaks in the past. Therefore, I use a moving window of the latest 80 quarterly observations of each data vintage to estimate the models. Aside from structural breaks the high inflation periods of the 70's and 80's influence the estimated inflation steady state which can bias the inflation forecasts of the late 80's and the 90's. A window of 80 observations gives at least the chance of a diminishing effect on the forecasts. The first sample for the FOMC meeting of March 1980 starts in 1960Q1 and ends in 1979Q4, the second sample for the FOMC meeting of April 1980 starts in 1960Q2 and ends in 1980Q1, and this goes on until the last sample for the FOMC meeting of December 2000 that starts in 1980Q4 and ends in 2000Q3.

I forecast annualized quarterly real output growth as measured by the GNP/GDP real growth rate, annualized quarterly inflation as measured by the GNP/GDP deflator and the federal funds rate. GDP data is first released about one month after the end of the quarter to which the data refer, the so-called advance release. These data series are then revised several times at the occasion of the preliminary release, final release, annual revisions and benchmark revisions. I follow Faust and Wright (2009) and use actual realized data as recorded in the data vintage that was released two quarters after the quarter to which the data refer to evaluate forecasting accuracy. For example, revised data for 1999Q1 is obtained by selecting the entry for 1999Q1 from the data vintage released in 1999Q3. Hence, I do not attempt to forecast annual and benchmark revisions, because the models cannot predict changes in data definitions. The revised data against which the accuracy of forecasts is judged will typically correspond to the final NIPA release.

DSGE model forecasts are compared to Greenbook projections and nonstructural forecasts from

Faust and Wright (2009). The dataset by Faust & Wright contains Greenbook nowcasts and forecasts up to five quarters ahead for all variables. However, I have spotted some differences between the Greenbook projections in the dataset by Faust & Wright and the Greenbook projection dataset that is published by the Federal Reserve Bank of Philadelphia.[6] The differences are neglectable for inflation and the interest rate. However, for GDP nowcasts and one quarter ahead forecasts the dataset by Faust & Wright understates the accuracy of the Greenbook projections quite a bit. A detailed comparison of the two datasets is contained in the appendix. For the evaluation of forecasts I have replaced the Greenbook projections from the Faust & Wright dataset with the Greenbook projections from the historical scanned Greenbook documents published by the Philadelphia Fed.[7]

4 Forecasting methodology

Computing recursive forecasts using structural models and real-time data vintages requires a sequence of steps that are explained in the following. First, the models need to be specified, solved and linked to the empirical data. Second, the data needs to be updated to the current vintage and parameters have to be estimated. Third, density and point forecasts are computed.

4.1 Model specification and solution

Each of the models consists of a number of linear or nonlinear equations that determine the dynamics of the endogenous variables. A number of structural shocks is included in each model. Any of the models $m = 1,...,4$ can be written as follows:

$$E_t\left[f_m(y_t^m, y_{t+1}^m, y_{t-1}^m, \varepsilon_t^m, \beta^m)\right] = 0 \qquad (1)$$

$$E(\varepsilon_t^m) = 0 \qquad (2)$$

$$E(\varepsilon_t^m \varepsilon_t^{m\prime}) = \Sigma_\varepsilon^m, \qquad (3)$$

where $E_t\left[f_m(.)\right]$ is a system of expectational difference equations, y_t^m is a vector of endogenous variables, ε_t^m a vector of exogenous stochastic shocks, β^m a vector of parameters and Σ_ε^m is the variance-covariance matrix of the exogenous shocks. The parameters and the variance-covariance matrix are

[6]The Federal Reserve Bank of Philadelphia has made available scanned pdf files of all Greenbook projections from 1966-2005: http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/pdf-data-set.cfm. Previously only Greenbook projections of four of the eight FOMC meetings per year were available as an Excel-file: http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/philadelphia-data-set.cfm. The update allowed a comparison of all Greenbook projections contained in the dataset by Faust and Wright (2009) with the original scanned documents. The assumed interest rate path that the Greenbook projections are conditioned on is also available from the website of the Federal Reserve Bank of Philadelphia: http://www.philadelphiafed.org/research-and-data/real-time-center/greenbook-data/gap-and-financial-data-set.cfm.

[7]Accordingly, I recalculate the root mean squared prediction errors (RMSE) of the nonstructural forecasts that are stated relative to Greenbook RMSEs by Faust and Wright (2009). I multiply the relative RMSEs with the RMSEs of the Greenbook projections by the Faust & Wright dataset and divide by the RMSEs of Greenbook projections of the Philadelphia Fed dataset.

either calibrated or estimated or a mixture of both.
A subset of the endogenous variables consists of empirically observable variables $y_t^{m,obs}$. If variables in the models are defined in percentage deviations from steady state then there is a subset of the equations that are so-called measurement equations $f_m^{obs}(.)$. These link the observable variables to the other endogenous variables through the inclusion of steady state values or steady state growth rates. Another possibility is that the observable variables are directly included in the structural equations of a model. This is the case for the FM model. For the EDO model, it is assumed that not all observable variables are measured exactly and therefore a set of nonstructural measurement shocks is added to the measurement equations.

The system of equations is solved using a conventional solution method for rational expectations models such as the technique of Blanchard and Kahn. Given the solution, the following state space representation of the system is derived:

$$
\begin{aligned}
y_t^{m,obs} &= \Gamma^m \bar{y}^m + \Gamma^m y_t^m + \varepsilon_t^{m,obs}, && (4)\\
y_t^m &= g_y^m(\beta^m) y_{t-1}^m + g_\varepsilon^m(\beta^m)\varepsilon_t^m, && (5)\\
E(\varepsilon_t^m \varepsilon_t^{m\prime}) &= \Sigma_\varepsilon^m && (6)
\end{aligned}
$$

The first equation summarizes the measurement equations and shows the link between observable variables and the endogenous model variables via steady state values or deterministic trends $\bar{y}^m$. The matrix Γ^m might include lots of zero entries as not all variables are directly linked to observables. The measurement errors $\varepsilon_t^{m,obs}$ are a subset of the shocks ε_t^m. The second equation constitutes the transition equations including the solution matrices g_y^m and g_ε^m that are nonlinear functions of the structural parameters β^m. The transition equations relate the endogenous variables to their own lags and the vector of exogenous shocks. The third equation denotes the variance-covariance matrix Σ_ε^m.

4.2 Estimation

Having solved the model and linked to the data, one needs to update the data to the relevant data vintage before estimating the model. Estimating DSGE models using Bayesian estimation has become a popular approach due to the combination of economic theory which is imposed on the priors and data fit summed up in the posterior estimates. A survey of the methodology is presented in An and Schorfheide (2007). Due to the nonlinearity in β^m the calculation of the likelihood is not straightforward. The Kalman filter is applied to the state space representation to set up the likelihood function.[8] Combining the likelihood with the priors yields the log posterior kernel $ln\mathscr{L}(\beta^m|y_1^{m,obs},...,y_t^{m,obs}) + lnp(\beta^m)$ that is maximized over β^m using numerical methods to compute the posterior mode. The posterior distribution of the parameters is a complicated nonlinear function

[8] I consider only unique stable solutions. If the Blanchard-Kahn conditions are violated I set the likelihood equal to zero.

of the structural parameters. The Metropolis-Hastings algorithm offers an efficient method to derive the posterior distribution via simulation. Details are provided for example in Schorfheide (2000). I compute 500000 draws from the Metropolis-Hastings algorithm and use the first 25000 of these to calibrate the scale such that an acceptance ratio of 0.3 is achieved. Another 25000 draws are disregarded as a burn in sample. The models are reestimated for the first data vintage of each year. Finally, the mean parameters can be computed from the posterior distribution of β^m.

4.3 Forecast computation

Having estimated the different models, forecasts for the horizons $h \in (0,1,2,3,4,5)$ are derived. First, a density forecast is computed and afterwards a point forecast is calculated as the mean of the density forecast. For each parameter a large number of values are drawn from the parameter's posterior distribution. For a random draw s a projection of the observable variables is derived by iterating over the solution matrix $g_y^m(\hat{\beta}^{m,s})$. At each iteration i in addition a vector of shocks $\varepsilon_i^{m,s}$ is drawn from a mean zero normal distribution where the variance is itself a random draw from the posterior distribution of the variance-covariance matrix:

$$\begin{aligned} y_{t+h}^{s,m,obs} &= \Gamma^m \hat{\bar{y}}^{m,s} + \Gamma^m g_y^m(\hat{\beta}^{m,s})^{h+1} y_{t-1}^m + \Gamma^m \sum_{i=0}^{h} g_\varepsilon^m(\hat{\beta}^{m,s})^{(h+1-i)} \varepsilon_i^{m,s} \quad (7) \\ \varepsilon_i^{m,s} &\sim N(0, \hat{\Sigma}_\varepsilon^{m,s}), \quad (8) \end{aligned}$$

where a hat on the structural parameters $\beta^{m,s}$, the variance covariance matrix $\Sigma_\varepsilon^{m,s}$ and the steady state values of observable variables $\bar{y}^{m,s}$ denotes that they are estimated. The reduced form solution matrices g_y^m and g_ε^m are functions of the estimated parameters and change over time as the models are reestimated. The procedure is repeated 10000 times ($s = 1, ..., 10000$) and finally the forecast density is given by the ordered set of forecast draws $y_{t+h}^{s,m,obs}$. The point forecast is given by the mean of the forecast density.

5 Forecast evaluation

Forecasting accuracy is evaluated in terms of root mean squared prediction errors (RMSE) for output growth, inflation and the federal funds rate. The performance of the DSGE models is compared to the Greenbook projections and the nonstructural forecasting models considered by Faust and Wright (2009). The results are shown in table 2. The first column shows the RMSE for the Greenbook and all other columns report the RMSEs of the different models relative to the Greenbook RMSE. Values less than one indicate that a model forecast is more accurate than the corresponding Greenbook projection. The last two columns report the relative RMSEs of the best and worst performing reduced form forecasting model from Faust and Wright (2009) for each horizon.

The first six rows in each table show forecasts based on the available data at the starting point of the forecast. The current state of the economy is not available in the data and therefore needs to be estimated. This nowcast is labeled as a forecast for horizon zero. As the data becomes available with a lag of one quarter, the results are labeled as "jump off -1". In practice, however, there are many data series that are available on a monthly, weekly or daily frequency that can be used to improve current-quarter estimates of GDP. Examples include industrial production, sales, unemployment, opinion surveys, interest rates and other financial prices. This data can be used to improve nowcasts and the Federal Reserve staff and many professional forecasters certainly make use of it.[9] To approximate the effect of using more information in nowcasting, I investigate the effect of using Greenbook nowcasts as a starting point for model-based forecasts regarding future quarters. The results are shown in the last five rows of each table and are labeled as "jump off 0".

I follow Faust and Wright (2009) in leaving out the period from 1980-1983 from the evaluation as this period was very volatile and might bias the assessment of forecasting accuracy for the whole sample. Therefore, the evaluation starts in 1984 so that RMSEs for output growth and inflation are directly comparable to table 2 in Faust and Wright (2009). Reported RMSEs are thus based on 123 forecasts from 1984 to 2000. I evaluate whether the difference of Greenbook RMSEs and model RMSEs is statistically significant based on the Diebold-Mariano statistic (Diebold and Mariano, 1995) using a symmetric loss function. Asymptotic p-values are computed using Newey-West standard errors with a lag-length of 10, covering a bit more than a year, to account for serial correlation of forecast errors.[10]

Results for inflation, output growth and the federal funds rate are very different from each other. For output growth the Greenbook nowcast is more precise than the model nowcasts. This was expected as the Fed can exploit more information about the current state of the economy. However, this precise estimate of the current state of the economy does translate only into a slightly superior forecasting performance one quarter ahead, but no more accurate forecasts than from different methods for higher horizons. The SW and EDO and models' forecasts dominate the Greenbook forecast from horizon 3 onwards. The DS model yields a similar forecasting accuracy as the Greenbook. Only the FM model is slightly less accurate than the Greenbook forecast for all horizons. If I include the Greenbook nowcast in the information set used to compute forecasts the results hardly change as quarterly output growth is not very persistent.

[9] Giannone et al. (2009) show how to incorporate such conjunctural analysis in structural models systematically. Employing such methods is, however, beyond the scope of this paper.

[10] The significance levels in columns *bestFW* and *worstFW* as copied from table 2 in Faust and Wright (2009) are only indicative. As described in section 3 I have used the Greenbook projections reported by the Philadelphia Fed, which are more reliable than the ones used by Faust & Wright. They indicate more accurate output growth nowcasts. Thus, it is likely that the output growth nowcasts in the last two columns are significantly different from the Greenbook nowcast. For the DSGE models differences in the output growth nowcast changed from insignificance to being significant on the 1% level when switching from the Greenbook projections reported by Faust & Wright to the ones reported by the Philadelphia Fed.

Table 2: Greenbook RMSE and relative RMSE of model forecasts: 1984-2000

(a) Output growth

horizon	GB	DS	FM	SW	EDO	best FW	worst FW
				jump off -1			
0	1.52	1.41●	1.32•	1.43●	1.40●	1.25	1.60
1	1.85	1.09	1.22	1.05	1.04	**0.99**	1.38
2	2.01	1.05	1.10	**0.92**	1.01	**0.95**	1.15
3	2.15	**0.99**	1.09	**0.86·**	**0.94**	**0.94**	1.12
4	2.08	1.00	1.05	**0.89**	**0.95**	**0.99**	1.11
5	2.08	1.02	1.06	**0.89**	**0.99**	**0.97**	1.09
				jump off 0			
1	1.85	1.08	1.19·	1.07	1.07	**0.96**	1.23
2	2.01	1.06	1.14	**0.92**	1.01	**0.90**	1.12
3	2.15	1.00	1.13	**0.86·**	**0.96**	**0.95**	1.18
4	2.08	1.01	1.08	**0.88**	**0.98**	**0.96**	1.09
5	2.08	1.03	1.08	**0.89·**	1.01	**0.98**	1.11

(b) Inflation

horizon	GB	DS	FM	SW	EDO	best FW	worst FW
				jump off -1			
0	0.70	1.51●	1.84●	1.49●	1.65●	1.32●	1.61●
1	0.80	1.57●	1.76●	1.42●	1.48●	1.20●	1.84●
2	0.82	1.35●	1.54●	1.27●	1.55●	1.14·	1.90●
3	0.94	1.16•	1.40•	1.17•	1.46●	1.02	1.82●
4	0.91	1.24•	1.74●	1.25●	1.43●	1.06	2.06●
5	1.15	1.22·	1.59●	1.22●	1.31●	**0.98**	1.81●
				jump off 0			
1	0.80	1.21•	1.58●	1.12	1.16·	1.19●	1.56●
2	0.82	1.25•	1.51●	1.18·	1.15	1.17	1.67●
3	0.94	1.24●	1.26•	1.20●	1.23•	1.03	1.64●
4	0.91	1.16●	1.46●	1.16•	1.22·	1.03	1.87●
5	1.15	1.13•	1.45•	1.18●	1.11	**0.96**	1.75●

(c) Federal Funds Rate

horizon	GB	DS	FM	SW	EDO	best FW	worst FW
				jump off -1			
0	0.11	6.38●	5.02●	4.95●	6.19●	-	-
1	0.52	2.00●	1.75●	1.76●	2.26●	-	-
2	0.94	1.42•	1.38•	1.30·	1.69●	-	-
3	1.29	1.14	1.21	1.06	1.50●	-	-
4	1.64	1.03	1.19·	**0.95**	1.38●	-	-
5	1.93	**0.96**	1.21·	**0.86**	1.29•	-	-
				jump off 0			
1	0.52	1.28·	1.24•	1.11	1.59●	-	-
2	0.94	1.13	1.03	1.02	1.48●	-	-
3	1.29	**0.99**	**0.97**	**0.92**	1.42●	-	-
4	1.64	**0.94**	1.00	**0.87**	1.36●	-	-
5	1.93	**0.89**	1.06	**0.82**	1.29●	-	-

Notes: GB: Greenbook; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO Model by Edge, Kiley & Laforte; best/worst FW: best/worst performing reduced from model for the specific horizon considered by Faust & Wright. Greenbook RMSEs are shown in levels while all other RMSEs are reported relative to the Greenbook. Values less than one are in bold and show that a forecast is more accurate than the one by the Greenbook. The symbols ●, •, ·, indicate that the relative RMSE is significantly different from one at the 1, 5, or 10% level, respectively.

Viewing the Greenbook as a best practice benchmark, one could be tempted to judge the forecasting ability of the structural models as very good at least for medium term horizons. However, one should keep in mind that quarterly output growth has little persistence and is difficult to forecast in general. The reported RMSEs in Faust and Wright (2009) show that none of their nonstructural forecasting methods is more accurate than a univariate autoregressive forecast. I find that only the SW model's forecasts are more precise than an autoregressive forecast from horizon 3 onwards. The forecasting accuracy of the EDO model is similar to the autoregressive forecast and the DS and FM forecasts are less precise. In addition, none of the models' RMSEs differs statistically significantly from the Greenbook RMSEs except for the nowcast. The difference in the forecasting accuracy of the models can be traced to the different modeling assumptions. The SW and EDO model have a richer economic structure than the DS and FM model and include a larger number of observable variables.

Greenbook inflation forecasts are more accurate than all structural as well as all nonstructural inflation forecasts. The structural forecasts have an accuracy in line with the accuracy range of the nonstructural forecasts. However, none of the DSGE models reaches the forecasting quality of the best nonstructural forecasts. The forecasting accuracy relative to the Greenbook forecasts improves with increasing horizons for all models. Adding the Greenbook nowcast to the information set of the models increases the forecasting accuracy quite a bit because of the persistence of inflation. The larger SW and EDO models perform better than the smaller DS and FM models. Still the model forecasts are less precise than the Greenbook projections. While it is not possible to forecast inflation with DSGE models as precisely as the Fed does, the forecasts are reasonable: with the exception of the FM model they are as good or better than a simple autoregressive forecast from horizon 3 onwards and for all horizons for the jump of 0 scenario.

Greenbook projections are conditioned on a hypothetical path of the federal funds rate. This path is not meant to be a forecast. Nevertheless, treating it as a forecast shows that its accuracy for short horizons is extremely high. Therefore, the Fed might have conditioned the projections on a policy path that is likely to be implemented in the next few quarters and it is reasonable to view this as a forecasting benchmark. Faust and Wright (2009) did not compute interest rate forecasts, so that I cannot compare the structural forecasts to forecasts from their nonstructural time series models.

Due to the extremely high accuracy of the Greenbook interest rate path in the short term, the structural forecasts perform much worse for horizons 0 to 2. Hence, for future research it might be of interest to compute DSGE model forecasts conditioned on the Greenbook interest rate path. There is the possibility that the short run accuracy of the assumed interest rate path increases the short-run forecasting accuracy for other variables as well. For medium term forecasts, however, the interest rate forecasting accuracy of the DS and SW models dominates the Greenbook path. Only the EDO forecasts are very imprecise as they underestimate the level of the interest rate many times. Taking the Greenbook nowcast as given, the forecasting accuracy of the models relative to the Greenbook increases because of the high persistence of the federal funds rate and there is no significant difference between model and central bank forecasts anymore.

Faust and Wright (2009) present a table showing the percentage of forecast periods in which the nonstructural forecasts are more accurate than the Greenbook. This metric is not as sensitive to outliers as the RMSEs. I compute accordant numbers for the DSGE models which are shown in the appendix. This metric leads to the same relative rankings of the forecasting models as in table 2.

5.1 Comparison with Bayesian VARs

To isolate the role played by information and the role played by the discipline of economic theory in determining forecasting accuracy I estimate several Bayesian VARs that contain the same variables as the DSGE models. In addition I estimate a large Bayesian VAR that uses all variables from the real-time dataset. In the Bayesian VARs all variables are treated symmetrically and therefore they incorporate no behavioral interpretations of parameters or equations.

Unrestricted VARs are heavily overparametrized and therefore not suitable for forecasting. To achieve a parsimonious parametrization I use a version of the Minnesota prior (see Doan et al., 1984). The Minnesota prior implies shrinking the parameters towards zero by setting a random walk or white noise prior. For the equations where inflation or the interest rate are the dependent variables, I set the prior on the first lag of the dependent variable to one and on all other lags to zero. Since, the DSGE models include output growth rather than the level of output, I set the the prior for the first lag of output growth to zero. The priors for all other variables are also set to zero as the DSGE models are based on the assumption that these variables are stationary. The prior variance of the parameters decreases with the lag length. The rationale for this assumption is that short lags contain more information about the dependent variables than long lags.

To control the overall tightness of the prior variance of the parameters, I use the version of the Minnesota prior by Bańbura et al. (2010) that increases the degree of shrinkage with increasing model size. They include a hyperparameter λ that controls the overall tightness of the prior distribution around the random walk or white noise prior. The three variable BVAR is treated as a benchmark model without any shrinkage and λ is set in a way to achieve the same in-sample fit for the larger BVARs as for the three variable BVAR. Thus, the tightness of the prior increases with the number of variables included in the BVAR. I estimate BVARs with the same three variables as in the DS and FM models (BVAR3), with the same seven variables as in the SW model (BVAR7) and the same 11 variables as in the EDO model (BVAR11). An additional large Bayesian VAR (BVARMax) includes all variables in each data vintage. The number of variables varies over the different vintages and the number of series in each vintage ranges from 47 to 80, with an average of 67. I use the same data transformations as in Faust and Wright (2009) to achieve stationary time series and accordingly set a white noise prior for all variables in the BVARMax except for inflation and the federal funds rate. Table 3 shows the RMSEs of the BVARs in comparison to the DSGE models.

For output growth the larger BVARs yield more accurate forecasts than the smaller BVARs. This resembles the findings that the SW and EDO output growth forecasts are more precise than the DS and FM forecasts. For DSGE models and BVARs of the same size there is no systematic difference in the forecasting performance. Therefore, imposing economic theory that is helpful in interpreting the forecast does not worsen forecasting performance. The BVARMax shows that using more information to forecast output growth does not increase the forecasting performance further in comparison to the SW and EDO model and the BVAR7 and BVAR11. The reason is most likely the low persistence of output growth which makes it impossible to increase the forecasting accuracy systematically over the one of a univariate forecasting model.

For inflation the BVAR3 performance is in between the performance of the DS and FM model. The SW model yields more accurate forecasts than the BVAR7, while the EDO model yields less accurate forecasts than the BVAR11. Overall, however, the differences between DSGE models and BVARs of the same size are small. Again, the restrictions imposed by economic theory do not worsen forecasting accuracy systematically in comparison to a purely data driven approach. For medium term horizons larger DSGE models and BVARs perform somewhat better than small DSGE models and BVARs. By including much more information as in the BVARMax the forecasting accuracy can be increased significantly. For medium term horizons the inflation forecasts from the BVARMax are not significantly different from the Greenbook projections anymore indicating that the Fed considers the information content of many variables when compiling the Greenbook inflation projections.

Regarding the federal funds rate forecasts, for short horizons it is apparent that the BVAR forecasts have a much higher accuracy than the corresponding DSGE model forecasts. The monetary policy rules in the DSGE models include only few variables and might be too simple. In contrast, the policy rules implicit in the BVARs contain four lags of the interest rate, output growth, the inflation rate and other variables. For the medium horizons and the jump off 0 scenario the BVARs yield RMSEs comparable to the DSGE counterparts. Only the EDO model performs much worse than the BVAR11. The reason is a frequent underestimation of the interest rate by the EDO model as discussed above. For short term horizons the forecasting accuracy of the BVARs increases with model size. The BVARMax yields the most accurate forecasts. This possibly reflects that the Fed studies a large variety of variables when deciding about the federal funds rate. When conditioning the forecasts on the Greenbook nowcast the advantage of larger over smaller models disappears.

Overall, the comparison shows that the economic structure of the DSGE models is no hindrance in achieving good forecasts. The DSGE models show a similar forecasting performance as their BVAR

Table 3: DSGE Model and BVAR RMSE: 1984-2000

(a) Output growth

horizon	GB	DS	FM	BVAR3	SW	BVAR7	EDO	BVAR11	BVARMax
					jump off -1				
0	1.52	1.41⬤	1.32●	1.41⬤	1.43⬤	1.34●	1.40⬤	1.33●	1.36●
1	1.85	1.09	1.22	1.23•	1.05	1.11	1.04	1.09	1.11
2	2.01	1.05	1.10	1.09	**0.92**	1.02	1.01	**0.97**	**0.94**
3	2.15	**0.99**	1.09	1.06	**0.86•**	**0.97**	**0.94**	**0.94**	**0.93**
4	2.08	1.00	1.05	1.00	**0.89**	0.99	**0.95**	**0.96**	**0.95**
5	2.08	1.02	1.06	1.02	**0.89**	0.92	**0.99**	**0.94**	**0.96**
					jump off 0				
1	1.85	1.08	1.19•	1.18●	1.07	1.04	1.07	1.01	1.09
2	2.01	1.06	1.12	1.03	**0.92**	1.00	1.01	**0.98**	**0.96**
3	2.15	1.00	1.13	1.10	**0.86•**	**0.97**	**0.96**	**0.93**	**0.95**
4	2.08	1.01	1.08	1.07	**0.88**	1.03	**0.98**	**0.96**	**0.97**
5	2.08	1.03	1.08	1.03	**0.89•**	**0.95**	1.01	**0.94**	**0.97**

(b) Inflation

horizon	GB	DS	FM	BVAR3	SW	BVAR7	EDO	BVAR11	BVARMax
					jump off -1				
0	0.70	1.51⬤	1.84⬤	1.49⬤	1.49⬤	1.41⬤	1.65⬤	1.47⬤	1.64⬤
1	0.80	1.57⬤	1.76⬤	1.50⬤	1.42⬤	1.49⬤	1.48⬤	1.47⬤	1.68⬤
2	0.82	1.35⬤	1.54⬤	1.37⬤	1.27⬤	1.41⬤	1.55⬤	1.39⬤	1.54⬤
3	0.94	1.16●	1.40●	1.31●	1.17●	1.41⬤	1.46⬤	1.30⬤	1.16
4	0.91	1.24●	1.74⬤	1.59●	1.25⬤	1.57⬤	1.43⬤	1.38⬤	1.31●
5	1.15	1.22•	1.59⬤	1.51•	1.21⬤	1.47⬤	1.31⬤	1.34⬤	1.32•
					jump off 0				
1	0.80	1.21●	1.58⬤	1.34⬤	1.12•	1.31⬤	1.16•	1.23⬤	1.23⬤
2	0.82	1.25●	1.51⬤	1.29⬤	1.18•	1.31⬤	1.15	1.30⬤	1.34⬤
3	0.94	1.24⬤	1.26●	1.25●	1.20⬤	1.42⬤	1.23●	1.30⬤	1.27⬤
4	0.91	1.16⬤	1.46⬤	1.34•	1.16●	1.49⬤	1.22•	1.31⬤	1.13
5	1.15	1.13●	1.45●	1.40●	1.18⬤	1.48⬤	1.11	1.29⬤	1.13

(c) Federal Funds Rate

horizon	GB	DS	FM	BVAR3	SW	BVAR7	EDO	BVAR11	BVARMax
					jump off -1				
0	0.11	6.38⬤	5.02⬤	4.96⬤	4.95⬤	4.45⬤	6.19⬤	4.04⬤	3.60⬤
1	0.52	2.00⬤	1.75⬤	1.88⬤	1.76⬤	1.69⬤	2.26⬤	1.61⬤	1.51⬤
2	0.94	1.42●	1.38●	1.29●	1.30•	1.28●	1.69⬤	1.24●	1.22●
3	1.29	1.14	1.21	1.10	1.06	1.12	1.50⬤	1.09	1.10
4	1.64	1.03	1.19•	1.04	**0.95**	1.07	1.38⬤	1.05	1.06
5	1.93	**0.96**	1.21•	1.00	**0.86**	1.03	1.29●	**0.99**	1.03
					jump off 0				
1	0.52	1.28•	1.24●	1.31●	1.11	1.11	1.59⬤	1.09	1.09●
2	0.94	1.13	1.03	1.14	1.02	1.02	1.48⬤	1.00	1.01
3	1.29	**0.99**	**0.97**	**0.97**	**0.92**	**0.97**	1.42⬤	**0.92**	**0.96**
4	1.64	**0.94**	1.00	**0.93**	**0.87**	**0.97**	1.36⬤	**0.92**	**0.96**
5	1.93	**0.89**	1.06	**0.91**	**0.82**	**0.96**	1.29⬤	**0.90**	**0.96**

Notes: GB: Greenbook; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; BVAR3: 3-variable BVAR; SW: Smets & Wouters; BVAR7: 7-variable BVAR; EDO: FRB/EDO Model by Edge, Kiley & Laforte; BVAR11: 11-variable BVAR; BVARMax: large BVAR using all variables in the real-time dataset.

counterparts. An increase in forecasting performance cannot be achieved by using more datadriven approaches, but by increasing the information set. On the one hand a larger number of variables increases the forecasting accuracy and even more importantly, conditioning on an accurate nowcast can increase the forecasting performance for higher horizons for persistent variables like inflation and the federal funds rate. Del Negro and Schorfheide (2004) propose to use DSGE models as priors for VARs. They show that the forecasting accuracy of these so-called DSGE-VARs improves relative to a VAR and partly to a BVAR with Minnesota priors. They advocate to use DSGE-VARs for forecasting until structural models are available that have the same forecasting performance. The forecasting results in table 2 show that at least the forecasting performance of the larger DSGE models is already good enough to be considered for forecasting exercises on its own.

5.2 Real-time vs. revised data

Most forecast evaluation studies are performed using revised data. To assess whether real-time data revisions matter, I conduct a pseudo real-time experiment by reestimating the models again on a moving window of observations, but this time only using revised data from the last data vintage. Not only the data used for estimation, but also the data realizations used for the forecast evaluation change. For the real-time results in tables 2 and 3 data released two quarters after the quarter to which the data refer were used as data realizations. In this section I use revised data realizations from the last data vintage. Therefore, even though the Greenbook projections remain unchanged, their RMSEs change, because of the difference in the data realizations. Table 4 shows the results.

The Greenbook RMSEs increase for output growth and decrease for inflation. The reason is that data realizations based on revised data are more volatile than those based on real-time data for output growth and less volatile for inflation. For example data realizations used for one step ahead predictions of output growth have a standard deviation of 2.11 for revised data and of 1.94 for real-time data. The corresponding realizations of inflation have a standard deviation of 0.93 for revised data and of 1.18 for real-time data. For the federal funds rate the Greenbook RMSEs remain unchanged as there are no data revisions for interest rate data. While the table only shows the RMSEs of the DSGE models relative to the Greenbook, one can easily compute the level of the RMSEs of the models which has increased uniformly for output growth and decreased for inflation. Similar results are documented in Faust and Wright (2009) for reduced form forecasting models. Hence, ignoring data revisions by using revised data gives a pessimistic view of the precision of output growth forecasts and an optimistic view of inflation forecasts. Interestingly, for the federal funds rate the forecasting performance of the DSGE models gets worse for the SW and EDO model when using revised data. This might reflect that models estimated on real-time data, i.e. the data that has actually been used by policy makers to set the interest rate, might yield a more realistic picture of actual interest rate setting

Table 4: Revised Data: Greenbook RMSE and relative RMSE of model forecasts: 1984-2000

(a) Output growth

horizon	GB	DS	FM	SW	EDO
			jump off -1		
0	1.76	1.23•	1.24·	1.42●	1.11
1	2.03	1.06	1.16	1.13	1.01
2	2.22	1.02	1.07	1.07	1.02
3	2.35	**0.96**	**0.99**	1.01	**0.99**
4	2.35	**0.95**	**0.96**	1.04	**0.96**
5	2.32	**0.99**	**0.99**	1.08	**0.99**
			jump off 0		
1	2.03	1.07	1.12	1.30•	1.03
2	2.22	1.01	1.12	1.08	**0.97**
3	2.35	**0.97**	1.03	1.00	**0.99**
4	2.35	**0.95**	1.00	1.03	**0.95**
5	2.32	**0.99**	**0.98**	1.05	1.02

(b) Inflation

horizon	GB	DS	FM	SW	EDO
			jump off -1		
0	0.60	1.36●	1.69●	1.35●	1.41●
1	0.70	1.35●	1.53●	1.41●	1.41●
2	0.78	1.11	1.37●	1.24•	1.33●
3	0.85	**0.99**	1.43•	1.26●	1.39●
4	0.96	1.18·	1.64●	1.39●	1.37●
5	1.04	1.11	1.67●	1.34•	1.30•
			jump off 0		
1	0.70	1.17•	1.39●	1.17•	1.24•
2	0.78	1.21•	1.41●	1.27●	1.31●
3	0.85	1.02	1.19	1.18·	1.20·
4	0.96	1.06	1.39•	1.24•	1.31•
5	1.04	1.10	1.52•	1.29●	1.25•

(c) Federal Funds Rate

horizon	GB	DS	FM	SW	EDO
			jump off -1		
0	0.11	6.40●	3.85●	5.24●	7.15●
1	0.52	2.06●	1.40•	1.98●	2.56●
2	0.94	1.46•	1.17	1.48•	1.88●
3	1.29	1.17	1.04	1.23	1.63●
4	1.64	1.04	1.01	1.09	1.48●
5	1.93	**0.96**	1.00	**0.98**	1.37●
			jump off 0		
1	0.52	1.26	**0.94**	1.14	1.62●
2	0.94	1.13	**0.83**	1.13	1.52●
3	1.29	1.00	**0.84**	1.05	1.43●
4	1.64	**0.95**	**0.88**	1.01	1.37●
5	1.93	**0.91**	**0.91**	**0.94**	1.30•

Notes: GB: Greenbook; DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO Model by Edge, Kiley & Laforte as found in the literature on monetary policy rule estimation (see e.g. Orphanides, 2001).

Regarding the relative performance of different forecasting models whether one uses real-time or revised data is less important. The ranking of the forecasting accuracy of the DSGE models is very similar to the real-time data experiment for output growth, inflation and the interest rate. Only the SW results are different. In the real-time data setting the SW model often had the lowest RMSE, while the performance of the SW model in the revised data setting is more in the middle of the accuracy range of the DSGE models. The forecasting performance of the DSGE models relative to the Greenbook increases for output growth and inflation when using revised data. This is not surprising as the models are reestimated on data that has the same characteristics as the data realizations, while the Greenbook projections are still based on real-time data, but are now evaluated against revised data with its different volatility.

6 Model averaging

The real-time forecast evaluation results in the previous section suggest that the forecasts from the SW model are more precise than those of the other considered DSGE models. The usage of revised data already showed that this results is not very robust. Table 5 shows for three subsamples the model with the lowest RMSE for the different forecast horizons and variables. There is no model that continuously performs better than all other models. Even for a given subsample the most precise forecasts for the different variables can be generated by different models. Different frictions in different models seem to be useful for forecasting specific variables in certain periods only, while other frictions are more important for other periods. The problem of instability in the performance and predictive content of different forecasting methods is well known (Rossi, 2012) and can for reduced form models be solved by computing combined forecasts. Forecasting accuracy usually increases (Timmermann, 2006). I will check in the following whether this also holds for forecast combinations of DSGE models.

I consider several methods to combine forecasts from the set of models: likelihood based weights, relative performance weights based on past RMSEs, a least squares estimator of weights, and non-parametric combination schemes (mean forecast, median forecast and weights based on model ranks reflecting past RMSEs). While many of these methods have been applied to nonstructural forecasts there are to my knowledge no applications to a suite of structural models. From a theoretical point of view likelihood based weights or weights estimated by least squares are appealing. In practice, these estimated weights have the disadvantage that they introduce estimation errors. In the applied literature simple combination schemes like equal-weighting of all models have widely been found to perform better than theoretically optimal combination methods (see e.g. Hsiao and Wan, 2010, for the disconnect of Monte Carlo simulation results and empirical results).

Table 5: Best performing forecasting models for three subsamples

(a) Output Growth

horizon	1984-1990	1990-1995	1995-2000
0	FM	EDO	FM
1	DS	EDO	FM
2	SW	EDO	FM
3	SW	EDO	FM
4	SW	SW	FM
5	SW	SW	FM

(b) Inflation

horizon	1984-1990	1990-1995	1995-2000
0	EDO	SW	FM
1	EDO	SW	FM
2	SW	SW	FM
3	SW	SW	FM
4	DS	SW	FM
5	DS	SW	FM

(c) Federal Funds Rate

horizon	1984-1990	1990-1995	1995-2000
0	EDO	SW	DS
1	SW	FM	DS
2	SW	FM	DS
3	SW	FM	DS
4	FM	SW	DS
5	SW	SW	DS

Notes: The table shows for each forecast horizon the best performing model in terms of RMSEs for three subsamples. DS: Del Negro & Schorfheide; FM: Fuhrer & Moore; SW: Smets & Wouters; EDO: FRB/EDO model by Edge, Kiley & Laforte.

6.1 Forecast combination methods

Here, I very shortly describe different methods to combine forecasts from the four DSGE models. A more detailed description can be found in the appendix. A natural way to weight different models in a Bayesian context is to use Bayesian model averaging. Posterior probability weights can be computed based on the marginal likelihood for each model. Unfortunately, posterior probability weights are not comparable for models that are estimated on a different number of time series. A second problem of the posterior probability weights is that over-parameterized models that have an extreme good in-sample fit, but a bad out-of-sample forecasting accuracy, are assigned high weights.

To circumvent these problems I use an out-of-sample weighting scheme based on predictive likelihoods as proposed by Eklund and Karlsson (2007) and Andersson and Karlsson (2007). The available data is split into a training sample used to estimate the models and a hold-out sample used to evaluate each model's forecasting performance. The forecasting performance is measured by the predictive likelihood, i.e. the marginal likelihood of the hold-out sample conditional on a specific model. To make the results comparable among models, only the three common variables output growth, inflation and the interest rate are considered for the computation of the predictive likelihood.

A second combination method is based on the assumption of a linear-in-weights model estimated with ordinary least squares (see Timmermann, 2006). I regress data realizations on the forecasts from the different models from previous data vintages separately for each variable. The resulting parameter estimates are the combination weights. The combination weights differ for different horizons and also for the three different variables. I omit an intercept term and restrict the weights to sum to one so that the weights can be interpreted as the fractions the specific models contribute to the weighted forecast. It also ensures that the combined forecast lies inside the range of the individual forecasts.

Thirdly, I consider weights based on past relative forecast performance. I use here weightings based on RMSEs of forecasts based on earlier data vintages. An additional modified version is to rank the RMSEs and base forecast weights on ranks rather than on the RMSEs directly to increase robustness.

Finally, the simplest method to compute a weighted forecast is to give equal weight to each model and simply compute the mean forecast of all models. From a theoretical point of view this approach is not preferable as the weights are purely subjective prior weights implicitly given by the choice of models. However, it has often been found that simple weighting schemes perform well (see e.g. Hsiao and Wan, 2010). A reason is that weight is given to several models instead of choosing one model that might perform poorly over the next quarters. Using the median rather than the mean forecast is another possibility.

6.2 Forecast Evaluation of combined forecasts

In table 6 I report RMSEs based on the six forecast weighting schemes described above. RMSEs are reported again relative to the Greenbook RMSEs. The second last column shows for comparison the relative RMSEs of the best single model as reported in table 2 and the last column shows the relative RMSEs of the best nonstructural model for each horizon as computed by Faust and Wright (2009). For all three variables it is apparent that the weighted forecasts have in general a higher accuracy than forecasts from single models.

For output growth the Greenbook nowcast is more accurate than all other forecasts, but for all other horizons the weighted model forecasts dominate the Greenbook forecast. The predictive likelihood (PL) weighting scheme is an exception with a forecasting quality not better, but still comparable to the Greenbook. There is not much of a difference between the accuracy of the other combination schemes. Most methods yield a similar forecasting accuracy as the best nonstructural forecasts and for medium horizon forecasts even dominate those.

For the inflation forecast weighted forecasts increase the forecasting accuracy compared to sin-

Table 6: Greenbook RMSE and relative RMSE of weighted model forecasts: 1984-2000

(a) Output growth

horizon	GB	PL	OLS	Median	Mean	RMSE	Rank	best M	best FW
				jump off -1					
0	1.52	1.41⬤	1.26●	1.26⬤	1.25⬤	1.24⬤	1.23⬤	1.32●	1.25
1	1.85	1.09	1.02	**0.99**	**0.99**	**0.98**	**0.99**	1.04	**0.99**
2	2.01	1.05	**0.99**	**0.92**	**0.92**	**0.92**	**0.91**	**0.92**	**0.95**
3	2.15	**0.99**	**0.89**	**0.90•**	**0.89•**	**0.89•**	**0.87●**	**0.86•**	**0.94**
4	2.08	1.00	**0.90**	**0.90**	**0.90**	**0.89**	**0.86•**	**0.89**	**0.99**
5	2.08	1.02	**0.91**	**0.93**	**0.92**	**0.92**	**0.89**	**0.89**	**0.97**
				jump off 0					
1	1.85	1.08	1.00	**0.97**	**0.97**	**0.97**	**0.98**	1.07	**0.96**
2	2.01	1.06	**0.94**	**0.91•**	**0.91**	**0.91•**	**0.91•**	**0.92**	**0.90**
3	2.15	1.00	**0.92**	**0.89•**	**0.89•**	**0.89•**	**0.88●**	**0.86•**	**0.95**
4	2.08	1.01	**0.92**	**0.91**	**0.90**	**0.89**	**0.89**	**0.88**	**0.96**
5	2.08	1.03	**0.95**	**0.93**	**0.92**	**0.92**	**0.94**	**0.89•**	**0.98**

(b) Inflation

horizon	GB	PL	OLS	Median	Mean	RMSE	Rank	best M	best FW
				jump off -1					
0	0.70	1.51⬤	1.68⬤	1.43⬤	1.45⬤	1.45⬤	1.45⬤	1.49⬤	1.32⬤
1	0.80	1.57⬤	1.59⬤	1.47⬤	1.44⬤	1.45⬤	1.47⬤	1.42⬤	1.20⬤
2	0.82	1.35⬤	1.37⬤	1.25●	1.22●	1.22●	1.25⬤	1.27⬤	1.14•
3	0.94	1.16●	1.25⬤	1.10	1.06	1.07	1.09	1.16●	1.02
4	0.91	1.24●	1.30●	1.17	1.11	1.11	1.15	1.24•	1.06
5	1.15	1.22•	1.23	1.14	1.08	1.09	1.12	1.22•	**0.98**
				jump off 0					
1	0.80	1.21●	1.22●	1.12	1.15•	1.14•	1.15•	1.12	1.19⬤
2	0.82	1.25●	1.27⬤	1.19•	1.16•	1.17•	1.17•	1.15	1.17
3	0.94	1.24⬤	1.26⬤	1.13●	1.08	1.09	1.10	1.20⬤	1.03
4	0.91	1.16⬤	1.16•	1.06	1.03	1.04	1.09	1.16●	1.03
5	1.15	1.13●	1.12	1.07	1.03	1.04	1.06	1.11	**0.96**

(c) Federal Funds Rate

horizon	GB	PL	OLS	Median	Mean	RMSE	Rank	best M	best FW
				jump off -1					
0	0.11	6.38⬤	5.42⬤	4.43⬤	4.07⬤	3.92⬤	3.62⬤	4.95⬤	-
1	0.52	2.00⬤	2.12⬤	1.63⬤	1.42●	1.41●	1.37●	1.75⬤	-
2	0.94	1.42●	1.52●	1.22	1.12	1.11	1.10	1.30•	-
3	1.29	1.14	1.34●	1.02	**0.97**	**0.97**	**0.99**	1.06	-
4	1.64	1.03	1.25●	**0.96**	**0.95**	**0.94**	**0.95**	**0.95**	-
5	1.93	**0.96**	1.17●	**0.92**	**0.93**	**0.91**	**0.91**	**0.86**	-
				jump off 0					
1	0.52	1.28•	1.52⬤	1.08	1.00	1.00	1.05	1.11	-
2	0.94	1.13	1.45●	**0.99**	**0.91**	**0.91**	**0.95**	1.02	-
3	1.29	**0.99**	1.27•	**0.90**	**0.88**	**0.86**	**0.92**	**0.92**	-
4	1.64	**0.94**	1.22•	**0.89**	**0.90**	**0.87**	**0.91**	**0.87**	-
5	1.93	**0.89**	1.17•	**0.88**	**0.88**	**0.87**	**0.89**	**0.82**	-

Notes: PL: Predictive Likelihood; OLS: Ordinay Least Squares; Median: Median forecast; Mean: Mean forecast; RMSE: weighted by inverse RMSE; Rank: weighted by inverse ranks; best M: best single model forecast; best FW: best performing atheoretical model for the specific horizon considered by Faust & Wright.

gle model forecasts, especially for medium term horizons. The weighting schemes can roughly be divided into two groups: the PL and OLS weighted forecasts are less precise than the Median, Mean, RMSE and Rank weighted forecasts. The simple Mean forecast is most accurate and its precision is comparable to the best nonstructural forecast and for medium horizons to the Greenbook projections. Forecasting accuracy relative to the Greenbook increases with increasing horizons for all weighting schemes. This shows again that structural forecasts are especially useful for medium term forecasts.

Combining the forecasts of all four DSGE models improves the forecasting quality also for the federal funds rate forecasts. While the Greenbook interest rate path is more accurate for horizons 0 to 2, the Mean, RMSE and Rank weighted forecasts are more precise for horizons 3 to 5. The relative forecasting accuracy improves with increasing horizons for all weighting schemes. Taking the Greenbook nowcast as given, the accuracy of all weighting schemes increases due to the high persistence of the interest rate. The Mean forecast is as precise as the Greenbook policy path for horizon 1 and dominates it for all other horizons.

Overall it turns out that model combination methods that give weight to several models perform well. They are a hedge against choosing inaccurate forecasts from one model only. Likelihood based weighting methods are preferable in theory, but do not work as well in practice. Differences in predictive likelihoods of different models are so high that most times all weight is given to a single model. Also estimated weights by least squares do not perform as good as simpler combination schemes: restricting the weights to sum to one leads to estimation problems so that in some cases weight is given only to one model.[11] The best forecasting performance is achieved by the Mean forecast and the RMSE and Rank based weighted forecasts. The differences between these are very small. Therefore, at this stage, one can conclude that a simple Mean forecast is the preferable method. It is very easy to compute as one needs no forecasts and realization from earlier data vintages to calculate model weights and it yields precise forecasts that are quite robust to outliers. Using a simple mean is also helpful for the structural interpretability of DSGE model forecasts. One can simply sum up the demand shock, supply shock and monetary policy shock contributions to the forecasts of the four different models and divide the overall contributions of these different shocks by the number of models.

7 Density Forecast Evaluation

Assuming a symmetric loss function, the accuracy of point forecasts can be easily compared by computing RMSEs. Evaluating density forecasts is less straightforward. The true density is never observed. Still one can compare the distribution of observed data with density forecasts to check

[11] In the appendix I report as an example model weights for forecasts derived from the last data vintage.

whether forecasts provide a realistic description of actual uncertainty.[12]

I use the following evaluation procedure: I split up the density forecasts into probability bands that each cover 5% of the probability mass. For each data realization I can check into which of the 20 probability bands of the accordant density forecast it falls. Doing this for all realization and the corresponding density forecasts, 5% of the realizations should be contained in each of the probability bands. Otherwise the density forecasts are not a good characterization of the distribution of the data realizations. In general, if one divides density forecasts into probability bands of equal coverage, data realisations should be uniformly distributed across all probability bands. This is the approach outlined in Diebold et al. (1998) and Diebold et al. (1999). More formally, it is based on the relationship between the data generating process and the sequence of density forecasts via probability integral transforms of the observed data with respect to the density forecasts. The probability integral transform (PIT) is the cumulative density function corresponding to the sequence of n density forecasts $\{p_t(y^{obs}_{t+h})\}^n_{t=1}$ evaluated at the corresponding observed data points $\{y^{obs}_{t+h}\}^n_{t=1}$:

$$z_t = \int_{-\infty}^{y^{obs}_{t+h}} p_t(u)du, \qquad \text{for} \quad t = 1, ..., n. \tag{9}$$

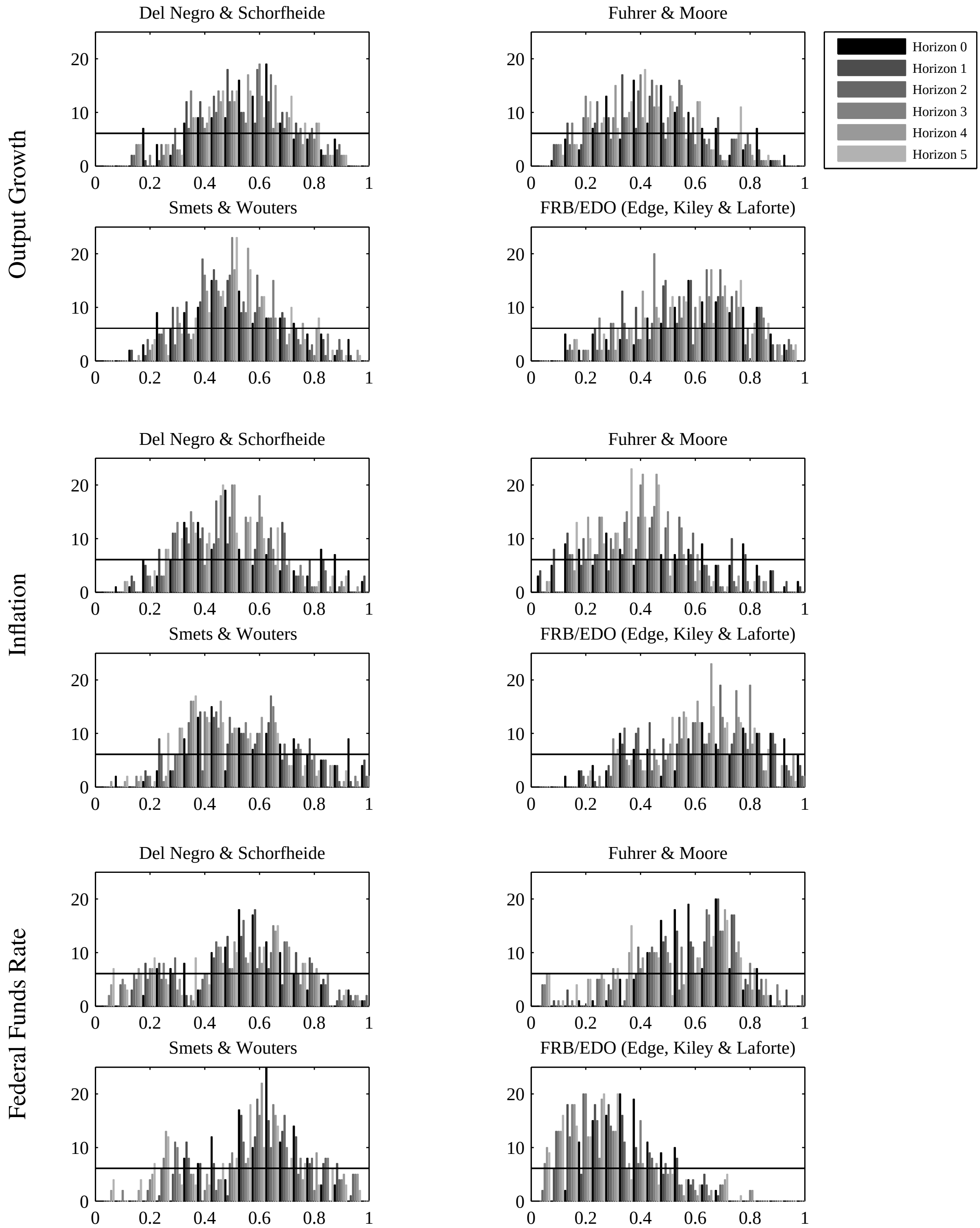
The PIT is the probability implied by the density forecast that a realized data point would be equal or less than what is actually observed. If the sequence of density forecasts is an accurate description of actual uncertainty, the sequence of PITs, $\{z_t\}^n_{t=1}$, should be distributed uniformly between zero and one. Figure 1 presents a visual assessment of the distribution of realized data points on the sequence of PITs that is represented as a histogram of 20 probability bands each covering 5%. There are $n = 123$ forecasts, so that there should be about 6 observations in each of the probability bands if the density forecasts are accurate. This is represented by the horizontal line. The bars in different grey shades reflect PITs for the different forecasting horizons.

The peak in the middle of the histograms of the output growth forecasts shows that these overestimate actual uncertainty. The histograms for inflation are somewhat closer to a uniform distribution, especially for the inflation nowcast. There is still a peak in the middle of the distributions, but the histogramms for some models cover the entire distribution including the tails. The density forecasts are imprecise for the federal funds rate. The tails are not covered, especially for short horizons, and thus uncertainty is overestimated by the density forecasts.[13] Gerard and Nimark (2008) give a plausible reason for the overestimation of actual uncertainty by DSGE models. The models impose tight restrictions on the data. If the data rejects these restrictions, large shocks are needed to fit the models

[12]While one can rank the accuracy of density forecasts from different models using the log score, I will check here whether DSGE models yield reasonable forecasts at all, i.e. I will look at the absolute rather than relative forecasting performance.

[13]In principle, there are tests available to formally check for a uniform distribution (Berkowitz, 2001). Unfortunately, the results have to be treated with high caution (see Elder et al., 2005; Gerard and Nimark, 2008). As the visual assessment has already shown clear evidence against a uniform distribution of the PITs, I do not use additional formal tests.

Figure 1: Evaluation of Structural Density Forecasts; 1984 - 2000



Notes: The figures show the distribution of realized data points on the density forecasts. The density forecasts are represented as probability bands each covering 5% of the density. The bars show how many of the realized observations fall in each of the probability bands. If the density forecast is an accurate description of actual uncertainty, than about six of the 123 observations should fall in each probability band.

to the data resulting in high shock uncertainty. Indeed, screening over all the forecasts for the different historical data vintages reveals that the models are suited to forecast during normal times. Given small or average exogenous shocks the models give a good view about how the economy will return back to steady state, but it is not possible to predict recessions and booms and the models require large exogenous shocks to capture these.

Herbst and Schorfheide (2012) also use PITs to evaluate density forecasts of a small New Keynesian model and the SW model. They find that for the SW model the density forecasts for output growth are too wide. The interest rate forecasts of both models are skewed as the federal funds rate was lower than predicted over the evaluation sample. The inflation forecasts of both models and the output forecasts of the small New Keynesian model perform well. The differences to my results might be due to differences in the sample and the evaluation data. Their estimation sample starts in 1984, while I use a rolling window of the most recent 80 quarterly observations. Their evaluation sample ranges from 1997 to 2004, while my evaluation sample starts in 1984 and ends in 2000. At least in the first half of my estimation and evaluation samples I estimate the models on relatively volatile data, but evaluate the forecasts on data that belong to the Great Moderation. This puts the DSGE models to a very hard test and might partly explain the overestimation of actual uncertainty. A possible solution would be to use models with time varying stochastic volatility and time varying Bayesian VARs.

The inclusion of the recent financial crisis could potentially lead to a reduction in the difference of predicted and actual uncertainty relative to my current results at least at the left tail of the forecast distribution.[14] I leave it for future research to assess the sensitivity of DSGE models' density forecast performance to the estimation and evaluation sample.

I have also checked the density forecasts of weighted forecasts. As all individual model forecasts overestimate actual uncertainty also the weighted density forecasts overestimate uncertainty as well.

8 Conclusion

Theory based DSGE models that are consistently derived from microeconomic optimization problems of households and firms have become the workhorse of modern monetary economics. Despite their stylized nature and their reliance on few equations they are widely used in academic work as well as at policy institutions. Computing out of sample forecasts is an ultimate test of the ability of this class of models to explain business cycles. In this paper, I have assessed the accuracy of point and density forecasts of four DSGE models using real-time data. Point forecasts of DSGE models are at least for medium term horizons surprisingly precise. A comparison with Bayesian VAR forecasts shows that

[14]Wieland and Wolters (2011) evaluate point forecasts of DSGE models for the recent financial crisis (and four other recessions) and find that the DSGE models are not able to forecast the intensity and lengths of such a deep recession. Thus, data realizations will be at left tail of density forecasts for output growth, inflation and the interest rate for 2008 and 2009.

the economic structure of the DSGE models that is helpful in interpreting the forecast does not worsen forecasting performance. An increase in forecasting performance cannot be achieved by using more datadriven approaches, but by increasing the information set. On the one hand a larger number of variables increases the forecasting accuracy and even more importantly, conditioning on an accurate nowcast can increase the forecasting performance for persistent variables like inflation and the federal funds rate. The Greenbook inflation projections are still more precise than the model forecasts. The good inflation forecasting results from a large Bayesian VAR show that the Fed might use a much larger information set than used by the DSGE models to achieve precise forecasts. For different subsamples, different DSGE models perform best. This instability can be overcome by combining forecasts from several models. Weighted forecasts increase the forecasting accuracy. Combination methods that give significant weight to several models are preferable over methods that aim to identify a single best model. The accuracy of a simple mean of model forecasts is hard to beat by other forecast weighting methods. While the point forecasts are surprisingly precise, density forecasts of DSGE model overestimate uncertainty. A reason might be the transition from a highly volatile estimation period to the Great Moderation evaluation period.

References

Adolfson, M., Andersson, M. K., Lindé, J., Villani, M., Vredin, A., 2007a. Modern forecasting models in action: improving macroeconomic analyses at central banks. International Journal of Central Banking 3(4), 111–144.

Adolfson, M., Laseén, S., Lindé, J., Villani, M., 2007b. Bayesian estimation of an open economy DSGE model with incomplete pass-through. Journal of International Economics 72(2), 481–511.

An, S., Schorfheide, F., 2007. Bayesian analysis of DSGE models. Econometric Reviews 26(2-4), 113–172.

Andersson, M. K., Karlsson, S., 2007. Bayesian forecast combination for VAR models, Sveriges Riksbank Working Paper No 216.

Bache, I. W., Jore, A. S., Mitchell, J., Vahey, S. P., 2011. Combining VAR and DSGE forecast densities. Journal of Economic Dynamics and Control 35(10), 1659–1670.

Bańbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. Journal of Applied Econometrics 25, 71–92.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics 19(4), 465–474.

Bernanke, B. S., Boivin, J., 2003. Monetary policy in a data-rich environment. Journal of Monetary Economics 50(3), 525–546.

Christoffel, K., Coenen, G., Warne, A., 2008. The New Area-Wide Model of the euro area - a micro-founded open-economy model for forecasting and policy analysis, European Central Bank Working Paper 944.

Christoffel, K., Coenen, G., Warne, A., 2011. Forecasting with DSGE models. In: Clements, M. P., Hendry, D. F. (Eds.), Oxford Handbook on Economic Forecasting. Oxford University Press, USA.

Del Negro, M., Giannoni, M. P., Schorfheide, F., 2013. Inflation in the Great Recession and New Keynesian models, Federal Reserve Bank of New York Staff Report No. 618.

Del Negro, M., Schorfheide, F., 2004. Priors from general equilibrium models for VARs. International Economic Review 45(2), 643–673.

Diebold, F. X., Gunther, T. A., Tay, A. S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39(4), 863–883.

Diebold, F. X., Hahn, J., Tay, A. S., 1999. Multivariate density forecast evaluation and calibration in financial risk management: High-frequency returns on foreign exchange. Review of Economics and Statistics 81(4), 661–673.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business & Economic Statistics 13, 253–263.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Reviews 3, 1 – 100.

Edge, R., Gürkaynak, R., 2010. How useful are estimated DSGE model forecasts for central bankers? Brookings Papers on Economic Activity 2, 209–244.

Edge, R. M., Kiley, M. T., Laforte, J.-P., 2007. Documentation of the research and statistics divisions estimated DSGE model of the U.S. economy: 2006 version, finance and Economics Discussion Series, Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington, D.C.: 2007-53.

Edge, R. M., Kiley, M. T., Laforte, J.-P., 2008. Natural rate measures in an estimated DSGE model of the U.S. economy. Journal of Economic Dynamics and Control 32, 2512–2535.

Edge, R. M., Kiley, M. T., Laforte, J.-P., 2010. A comparison of forecast performance between federal reserve staff forecasts, simple reduced form models and a DSGE model. Journal of Applied Econometrics 25(4), 720–754.

Eklund, J., Karlsson, S., 2007. Forecast combination and model averaging using predictive measures. Econometric Reviews 26(2-4), 329–363.

Elder, R., Kapetanios, G., Taylor, T., Yates, T., 2005. Assessing the MPC's fan charts. Bank of England Quarterly Bulletin Autumn, 326–345.

Faust, J., Wright, J. H., 2009. Comparing Greenbook and reduced form forecasts using a large realtime dataset. Journal of Business and Economic Statistics 27(4), 468–479.

Forni, M., Hallin, M., Lippi, M., Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the Euro area? Journal of Monetary Economics 50, 1243–1255.

Fuhrer, J. C., 1997. Inflation/output variance trade-offs and optimal monetary policy. Journal of Money, Credit and Banking 29(2), 214–234.

Fuhrer, J. C., Moore, G., 1995a. Inflation persistence. The Quarterly Journal of Economics 110(1), 127–159.

Fuhrer, J. C., Moore, G., 1995b. Monetary policy trade-offs and the correlation between nominal interest rates and real output. The American Economic Review 85(1), 219–239.

Gerard, H., Nimark, K., 2008. Combing multivariate density forecasts using predictive criteria, Research Discussion Paper 2008-2, Reserve Bank of Australia.

Geweke, J., Amisano, G., 2012. Prediction with misspecified models. American Economic Review 102(3), 482–486.

Giannone, D., Monti, F., Reichlin, L., 2009. Incorporating conjunctural analysis in structural models. In: Wieland, V. (Ed.), The Science and Practice of Monetary Policy Today. Springer Science, pp. 41–57.

Goodfriend, M., King, R. G., 1997. The new neoclassical synthesis and the role of monetary policy. In: Bernanke, B. S., Rotemberg, J. J. (Eds.), National Bureau of Economic Research Macroeconomics Annual 1997. MIT Press, Cambridge, MA.

Herbst, E., Schorfheide, F., 2012. Evaluating DSGE model forecasts of comovements. Journal of Econometrics 171(2), 152–166.

Hsiao, C., Wan, S. K., 2010. Is there an optimal forecast combination?, Working Paper University of Southern California.

Ingram, B., Whiteman, C., 1994. Supplanting the Minnesota prior- forecasting macroeconomic time series using real business cycle model priors. Journal of Monetary Economics 49(4), 11311159.

Marcellino, M., Stock, J., Watson, M., 2003. Macroeconomic forecasting in the Euro area: Country-specific versus area-wide information. European Economic Review 47, 1–18.

Murchison, S., Rennison, A., 2006. ToTEM: The Bank of Canada's new quarterly projection model, Technical Reports 97, Bank of Canada.

Orphanides, A., 2001. Monetary policy rules based on real-time data. American Economic Review 91(4), 964–985.

Rossi, B., 2012. Advances in forecasting under instability. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2. Elsevier Publications.

Rotemberg, J. J., Woodford, M., 1997. An optimization-based econometric framework for the evaluation of monetary policy, in B. Bernanke and J. Rotemberg, (eds.), NBER Macroeconomics Annual, The MIT Press.

Schorfheide, F., 2000. Loss function-based evaluation of DSGE models. Journal of Applied Econometrics 15, 645–670.

Sims, C. A., 2002. The role of models and probabilities in the monetary policy process. Brookings Papers on Economic Activity 2, 1–40.

Smets, F., Christoffel, K., Coenen, G., Motto, R., Rostagno, M., 2010. DSGE models and their use at the ECB. Journal of the Spanish Economic Society 1(1), 51–65.

Smets, F., Wouters, R., 2004. Forecasting with a bayesian DSGE model: An application to the Euro area. Journal of Common Market Studies 42(4), 841–867.

Smets, F., Wouters, R., 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. The American Economic Review 97(3), 586–606.

Stock, J., Watson, M., 2002. Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association 97, 1167–1179.

Timmermann, A., 2006. Forecast combinations. In: Elliott, G., Granger, C. W. J., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Amsterdam: North Holland, pp. 135–196.

Waggoner, D. F., Zha, T., 2010. Confronting model misspecification in macroeconomics, manuscript, Federal Reserve Bank of Atlanta.

Wang, M.-C., 2009. Comparing the DSGE model with the factor model: An out-of-sample forecasting experiment. Journal of Forecasting 28(2), 167–182.

Wieland, V., Wolters, M. H., 2011. The diversity of forecasts from macroeconomic models of the U.S. economy. Economic Theory 47(2-3), 247–292.

Wieland, V., Wolters, M. H., 2012. Forecasting and policy making. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. Vol. 2. Elsevier Publications.

Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.