The Diversity of Forecasts from Macroeconomic Models of the U.S. Economy*

Volker Wieland
Goethe University Frankfurt, CEPR and CFS†

Maik H. Wolters
Goethe University Frankfurt

February 17, 2010

Abstract

This paper investigates the accuracy and heterogeneity of output growth and inflation forecasts during the current and the four preceding NBER-dated U.S. recessions. We generate forecasts from six different models of the U.S. economy and compare them to professional forecasts from the Federal Reserve's Greenbook and the Survey of Professional Forecasters (SPF). The model parameters and model forecasts are derived from historical data vintages so as to ensure comparability to historical forecasts by professionals. The mean model forecast comes surprisingly close to the mean SPF and Greenbook forecasts in terms of accuracy even though the models only make use of a small number of data series. Model forecasts compare particularly well to professional forecasts at a horizon of three to four quarters and during recoveries. The extent of forecast heterogeneity is similar for model and professional forecasts but varies substantially over time. Thus, forecast heterogeneity constitutes a potentially important source of economic fluctuations. While the particular reasons for diversity in professional forecasts are not observable, the diversity in model forecasts can be traced to different modeling assumptions, information sets and parameter estimates.

Keywords: forecasting, business cycles, heterogenous beliefs, forecast distribution, model uncertainty, Bayesian estimation

JEL-Codes: C53, D84, E31, E32, E37,

*We are grateful to Mordecai Kurz and participants at the Stanford Institute for Theoretical Economics Workshop on Diverse Beliefs for very helpful comments on this research.

†Corresponding Author: Volker Wieland; Mailing address: Goethe University of Frankfurt, Grueneburgplatz 1, Goethe Universitaet, House of Finance, 60323 Frankfurt am Main, Germany, E-mail: wieland@wiwi.uni-frankfurt.de. Other Author: Maik Wolters; Mailing address: Grueneburgplatz 1, House of Finance, 60323 Frankfurt am Main, Germany.

1 Introduction

Recent empirical studies have documented substantial variations in the accuracy and heterogeneity of professional forecasts of aggregate output and inflation (cf. Kurz, Jin and Motolese (2003, 2005), Giordani and Söderlind (2003), Kurz (2009) and Capistran and Timmermann (2009)). At the same time, theoretical research has emphasized that expectational heterogeneity itself can be an important propagation mechanism for economic fluctuations and a driving force for asset price dynamics. Theories of heterogeneous expectations and endogenous fluctuations have been advanced, for example, in Kurz (1994a, 1994b, 1996, 1997a,1997b, 2008), Brock and Hommes (1998), Kurz, Jin, and Motolese (2005), Chiarella, Dieci, and He (2007), Branch and McGough (2009), Branch and Evans (2009) and De Grauwe (2009).

Forecast heterogeneity arises for a number of different reasons. Invariably, forecasters need some type of forecast-generating framework. Such a framework may constitute of a fully developed economic structure, a non-structural collection of statistical relationships or a simple rule-of-thumb. Thus, important sources of belief heterogeneity are the particular modeling assumptions embedded in this forecasting framework. Another source of heterogeneity lies in the information used by the forecaster. The information set may differ in terms of the number of economic aggregates or prices for which the forecaster collects data as well as the timeliness of the data vintage used. This dataset is needed to estimate the state of the economy as well as any parameters that form part of the forecasting framework.

Unfortunately, the modeling assumptions, information sets and parameter estimates underlying the forecasts reported by the Fed or professionals participating in the SPF or Bluechip surveys are not publicly available and their impact on the precision and heterogeneity of those forecasts cannot be studied. Instead, this paper investigates the precision and heterogeneity of output growth and inflation forecasts from six different models of the U.S. economy and compares them to forecasts from the Federal Reserve's Greenbook and the Survey of Professional Forecasters. This comparison is conducted for successive quarter-by-quarter forecasts up to four quarters into the future during the five most recent recessions of the U.S. economy as dated by the NBER. We focus on periods around recessions because downturns and recoveries pose the greatest challenge for economic forecasters, and arguably expectational heterogeneity may itself play role in these shifts in economic activity.

The six macroeconomic models considered in this paper include three small-scale New-Keynesian models, which differ in terms of structural assumptions, a non-structural Bayesian VAR model, as well as two medium-scale New-Keynesian dynamic stochastic general equilibrium (DSGE) models of the type currently used in some of the world's leading central banks. The four small models are fit to three key macroeconomic time series: real GDP growth, inflation measured by the GDP deflator and the federal funds rate. The two medium-scale models are estimated on data for 7 and 11 variables, respectively. Variables used in these models include consumption, investment, wages and hours worked. The largest model even accounts for the breakdowns in durables versus non-durables and services consumption, residential versus business investment as well as the related deflator series. We consider each of the six macroeconomic models as a reasonable forecast-generator that could also be used in a professional context. Although, the five structural models all embody the popular modeling assumption of homogenous rational expectations, the different forecasts generated from the models can be used to compute estimates of model-based forecast heterogeneity. The rationale for choosing these particular models is discussed in more detail in the next section.

To be able to create model-based forecasts that are comparable to the historical Greenbook and SPF forecasts, we have to put them on a similar footing in terms of the data vintage used for parameter estimation and initial conditions. Thus, we have created a large data set that contains all the historical quarterly vintages of the 11 time series used in the largest model. At each point in time, we re-estimate all the model parameters on the basis of the data vintage that was available at that exact point in time. Then, given the state of the economy as measured by this data vintage we compute an estimate of the current state of the economy, that is the so-called *nowcast* as well as forecasts up to four quarters into the future. Then, we assess forecast precision relative to the revised data that became available during the subsequent quarters for the dates to which the forecasts apply. We conduct this assessment for the recessions of the U.S. economy in 2008/09, 2001, 1990/91, 1981/82 and 1980. Typically, the forecasting exercise is initiated 4 quarters prior to the trough determined by the NBER Business Cycle Dating Committee and ends 4 quarters after the trough.[1]

The approach taken in this paper is novel and breaks new ground in several respects. We are not aware of any comparable assessment of the forecasting accuracy of multiple structural macroeconomic models based on historical data vintages. An innovative recent paper by Faust and Wright (2009) studies the performance of forecasts from nonstructural time series models on the basis of real-time data similar to Bernanke and Boivin (2003). Recently, Edge, Kiley, and Laforte (2009) have provided an assessment of the real-time forecasting performance of a single structural model. Furthermore, to our knowledge there has never been an attempt to quantify the heterogeneity of model-based forecasts and to compare to survey forecasts in order to learn more about the extent, dynamics and sources of forecast heterogeneity.

We obtain a number of interesting findings with regard to the relative accuracy of model-based and professional forecasts as well as the extent and dynamics of forecast diversity. The mean model forecast comes surprisingly close to the mean SPF and Greenbook forecasts in terms of accuracy even though the models only make use of a small number of data series. Model forecasts compare particularly well to professional forecasts at a horizon of three to four quarters and during recoveries. The

[1]Exceptions are the 1980 and 2008/9 recessions. In first case, we start only 2 quarters prior to the peak because of data availability. In the second case, the trough is not yet determined. We start in 2007Q4 and end in 2009Q3.

extent of forecast heterogeneity is similar for model and professional forecasts but varies substantially over time. Thus, forecast heterogeneity constitutes a potentially important source of economic fluctuations. While the particular reasons for diversity in professional forecasts are not observable, the diversity in model forecasts can be traced to different modeling assumptions, information sets and parameter estimates.

The remainder of this paper proceeds as follows. Section 2 summarizes the most important features of the different macroeconomic models that we use to compute forecasts. Section 3 describes the estimation and forecasting methodology. Section 4 provides an illustrative example by forecasting the 2001 recession. The difference between model-based and professional nowcasts and their impact on forecasting performance in the current recession are demonstrated in section 5. Section 6 provides a comparison of forecast accuracy of model and professional forecasts. The extent and dynamics of forecast heterogeneity is studied systematically in section 7. Section 8 summarizes our findings and concludes.

2 Forecasting Models

In evaluating the performance of model-based forecasts, we consider six different models of the U.S. economy. Five of these models are structural New-Keynesian macroeconomic models, while the sixth model is a Bayesian VAR model. The latter model is representative of simple vector autoregression models that are often used to summarize macroeconomic dynamics without imposing strong theoretical restrictions and constitutes a useful empirical benchmark for comparison.

The structural models are chosen to reflect the state of the art in modern macroeconomics in academia and central banks. The New-Keynesian model as laid out by Rotemberg and Woodford (1997) and Goodfriend and King (1997) and developed in detail in Woodford (2003) and Walsh (2003) has quickly become the principal workhorse model in monetary economics[2]. This model is based on consistent microeconomic foundations in terms of the optimizing and forward-looking behavior of representative households and firms. It incorporates restrictions in terms of monopolistic competition and price rigidity that ensure important interactions between nominal and real economic variables. We consider two empirical implementations of this model. The first specification is taken from Del Negro and Schorfheide (2004). These authors use a Bayesian estimation methodology to fit the model to output, inflation and interest rate data. In the following, it is referred to as the *NK-DS model*. The second specification differs only in terms of the restrictions on how particular economic shocks enter the model. It is estimated in the present paper with the same Bayesian methodology as in Del Negro and Schorfheide (2004) and termed the *NK-WW model*.

[2]For recent discussions of the application of the New-Keynesian approach in practical monetary policy see Wieland (2009).

The above-mentioned New-Keynesian model is a small dynamic stochastic general equilibrium (DSGE) model with rational expectations. Such models have recently been criticized for assuming the existence of a representative household and the use of rational expectations. For this reason we consider also an earlier-generation New-Keynesian model estimated by Fuhrer (1997) and referred to as the *NK-Fu model* in our analysis. This model still accounts for rational, forward-looking behavior by market participants but does not impose the specific restrictions that would follow from a consistent derivation from microeconomic foundations with a representative optimizing household and firm. Fuhrer (1997) uses traditional maximum likelihood estimation to parameterize the model and we follow the same approach in re-estimating this model in the present paper. The difference in estimation methodology also constitutes a potentially interesting source of forecast heterogeneity.

Christiano, Eichenbaum and Evans (2005) further developed the New-Keynesian DSGE modeling approach and showed how to build medium-scale models that can fit a significant number of important empirical regularities of the U.S. economy. To do so they introduce additional dimensions for optimizing behavior as well as additional economic frictions. Such medium-scale models include physical capital in the production function and account for endogenous capital formation. Labor supply is modeled explicitly. Nominal frictions include sticky prices and wages and inflation and wage indexation. Real frictions include consumption habit formation, investment adjustment costs and variable capital utilization.

Building on Christiano et al (2005), Smets and Wouters (2003, 2007) introduced a complete set of economic shocks and showed how to use Bayesian methods for fitting such models to observed macroeconomic time series. We generate forecasts from a version of this model estimated with Bayesian methods and refer to it as the *CEE-SW model* in the following.

DSGE modeling rapidly gained in popularity around the world and researchers in academia and many central banks estimated larger and more sophisticated DSGE models for their respective countries. The fifth structural model in our forecasting pool is a version of the new DSGE model developed at the Federal Reserve by Edge, Kiley, and Laforte (2008). Following these authors we refer to it as the *FRB-EDO model*.

In additional to to the structural models we include a VAR model with four lags on output growth, inflation and the federal funds rate. Since all variables are treated symmetrically, the VAR model incorporates no behavioral interpretations of parameters or equations. Unrestricted VAR models are heavily over-parameterized and therefore not suitable for forecasting. Thus, we use a Bayesian approach with so-called Minnesota prior (see Doan, Litterman, and Sims, 1984) to shrink the parameters towards zero and render the VAR model more effective in forecasting. It is referred to as the *BVAR-WW model* in the following.

While the NK-DS, NK-WW, NK-Fu and BVAR-WW models are estimated on just three key macroeconomic variables, namely output growth, inflation and interest rates, the two larger medium-scale models are fit to 7 and 11 economic time series, respectively. The CEE-SW model is estimated on real GDP growth, inflation as measured by the GDP deflator, the federal funds rate, wages, hours worked, consumption and investment. The FRB-EDO model allows for further disaggregation. It features two production sectors, which differ in their pace of technological progress. This structure can capture the different growth rates and relative prices observed in the data. Accordingly, the expenditure site 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 estimated on eleven empirical 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.

Table 1 summarizes the most important features of the six models used for computing forecasts. Further details are provided in appendix A2 at the end of this paper.

Table 1: Model Overview

Name/Reference	Short Name	Type	Observable Variables	Original Sample
Del Negro and Schorfheide (2004)	NK-DS	standard 3-equation New Keynesian model with Calvo contracts and forward looking IS-equation	3: output growth, inflation, interest rate	1955Q3-2001Q3
New Keynesian Model estimated in this paper	NK-WW	standard 3-equation New Keynesian model with mark-up and preference shocks	3: output growth, inflation, interest rate	1960Q1-1979Q4
Fuhrer (1997)	NK-Fu	small-scale closed economy New-Keynesian model with relative real wage contracts and backward looking IS curve	3: output growth, inflation, interest rate	1966Q1-1994Q1
Christiano, Eichenbaum, and Evans (2005) as estimated in Smets and Wouters (2007)	CEE-SW	medium-scale closed economy DSGE-model of the type used by policy institutions	7: output growth, consumption growth, investment growth, inflation, wages, hours, interest rate	1966Q1-2004Q4
Edge et al (2008)	FRB-EDO	medium-scale closed economy DSGE-model developed at he Federal Reserve. Two sectors with different technology growth rates	11: output growth, inflation, interest rate, consumption of non-durables and services, consumption of durables, residential investment, business investment, hours, wages, inflation for consumer nondurables and services, inflation for consumer durables	1984Q1-2004Q4
Bayesian VAR estimated in this paper	BVAR-WW	Bayesian VAR with 4 lags and Minnesota priors	3: output growth, inflation, interest rate	1960Q1-1979Q4

3 Forecasting Methodology

This section demonstrates how the forecasts using structural macroeconomic models are computed. Three aspects are best distinguished and discussed separately, the model specification and solution, the estimation of the model parameters, and then the actual sequence of steps that need to be accomplished to generate quarter-by-quarter forecasts.

Model specification and solution. The simple New-Keynesian model estimated originally by Del Negro and Schorfheide (2004) serves as a good example. Del Negro and Schorfheide (2004) study the log-linearized approximation to the original nonlinear DSGE model. It consists of three linear equations: a New-Keynesian IS equation derived from the household's intertemporal first-order condition, a New-Keynesian Phillips curve equation derived from the price-setting problem of the firm under monopolistic competition faced with price rigidities, and the interest rate rule of the central bank:

$$x_t = E_t x_{t+1} - \tau^{-1}(R_t - E_t \pi_{t+1}) + (1-\rho_g) g_t + \rho_z \tau^{-1} z_t \quad (1)$$

$$\pi_t = \frac{\gamma}{r^*} E_t \pi_{t+1} + \kappa (x_t - g_t) \quad (2)$$

$$R_t = \rho_R R_{t-1} + (1-\rho_R)(\psi_1 \pi_t + \psi_2 x_t) + \varepsilon_{R,t} \quad (3)$$

Here, variables are defined as percentage deviations from their steady state. x_t denotes output, π_t inflation and R_t the federal funds rate. g_t is a government spending shock and z_t a technology shock. Both shocks follow an AR(1) process that is not shown explicitly. The monetary policy shock $\varepsilon_{R,t}$ is iid-normally distributed. $(\tau, \rho_g, \rho_z, \gamma, r^* m, \kappa, \rho_R, \psi_1, \psi_2)$ represent model parameters that need to be estimated.

The model is connected with the available data by adding measurement equations that link the model variables to observable quarterly output growth, quarterly inflation, and the quarterly federal funds rate:

$$YGR_t = \ln \gamma + \Delta x_t + z_t \quad (4)$$

$$INFL_t = \ln \pi^* + \pi_t \quad (5)$$

$$INT_t = \ln r^* + \ln \pi^* + R_t \quad (6)$$

The system of linear expectational difference equations that comprises model and measurement equations is then solved using a conventional solution method such as the technique of Blanchard and

Kahn and the state space representation of the system is derived:

$$y_t^{obs} = y(\theta) + \lambda + y_t^s, \tag{7}$$

$$y_t = g_y(\theta) y_{t-1} + g_u(\theta) u_t, \tag{8}$$

$$E(u_t u_t') = Q(\theta), \tag{9}$$

Here, the first equation summarizes the measurement equations, the second equation constitutes the transition equation and the third equation denotes the variance-covariance matrix Q. θ refers to the vector of structural parameters. Table 2 provides a summary linking the variables and parameters in the state space representation to those in the model and the measurement equations.

Table 2: State Space Representation and Model Equations

structural parameters	$\theta = (\tau, \rho_g, \rho_z, \gamma, r^* m, \kappa, \rho_R, \psi_1, \psi_2)$
observable variables	$y_t^{obs} = [\ YGR_t \quad INFL_t \quad INT_t\ ]'$
steady state	$y(\theta) = [\ 0 \quad \ln \pi^* \quad \ln r^* + \ln \pi^*\ ]'$
deterministic trend	$\lambda = [\ \ln \gamma \quad 0 \quad 0\ ]'$
subset of endogenous variables	$y_t^s = [\ \Delta x_t + z_t \quad \pi_t \quad R_t\ ]'$
endogenous variables	$y_t = [\ x_t \quad R_t \quad \pi_t \quad g_t \quad z_t\ ]'$
shocks	$u_t = [\ \varepsilon_{R,t} \quad \varepsilon_{z,t} \quad \varepsilon_{g,t}\ ]'$

The observable variables y_t^{obs} that are defined by the measurement equations are functions of the stationary steady state $y(\theta)$, of a subset of the endogenous variables expressed in deviations from steady state, y_t^s, and of the deterministic trend λ. The transition equation comprises the three decision rules. Its parameters are given by the two solution matrices g_y and g_u which are nonlinear functions of the structural parameters θ. Thus, the transition equations relate the endogenous variables y_t to lags of themselves and the vector of exogenous shocks u_t. Since, the measurement equations include the deterministic growth path caused by labor-augmenting technological progress no separate de-trending of the data is necessary.

Model Estimation. Whenever possible, we estimate the models using the same techniques as the original authors. The model by Fuhrer (1997) is estimated using maximum likelihood techniques while the NK-DS, CEE-SW and FRB-EDO models are estimated using a Bayesian methodology. We also use Bayesian methods to estimate the NK-WW and BVAR-WW models. Maximum likelihood estimation maximizes the likelihood of the model, while Bayesian estimation combines the likelihood with prior beliefs obtained from economic theory, microeconomic data or previous macro studies. An extensive survey of the methodology is presented in An and Schorfheide (2007).

Because of the nonlinearity in the structural parameters, θ, the calculation of the likelihood is not straightforward. The Kalman filter is applied to the state space representation to set up the likelihood function (see e.g. Hamilton, 1994, chapter 13.4)[3]. Since the models considered here are stationary we can initialize the Kalman Filter using the unconditional distribution of the state variables. Combining the likelihood with the priors yields the log posterior kernel $ln\mathscr{L}(\theta|Y^T) + lnp(\theta)$ that is maximized over θ using numerical methods so as to obtain the posterior mode. We use the posterior mode to generate point forecasts. As a check we simulated the posterior distribution in some cases using the Metropolis-Hastings-Algorithm and compare point forecasts obtained from the posterior mean and posterior mode. Since the point forecasts were quite similar we relied therefore on the posterior mode for forecast generation in light of computational complexity of a large number of model re-estimations.

To estimate the Bayesian VAR we use a version of the Minnesota prior (see Doan et al, 1984) to shrink the parameters towards zero. The Minnesota prior assumes that the vector of time series is well-described as a collection of independent random walks. In our case we deal with growth rates or stationary time series and therefore put a prior assumption of a zero coefficient on the first lag of the dependent variable instead of a coefficient of unity. All parameters are assumed to be normally distributed with mean zero. The prior variance of the parameters decreases with the lag length.

Forecasting. For a given date, we estimate each of the models on the basis of the most recent data vintage that would have been available at that time. Thus, data vintages are identical across models and change quarter-by-quarter as in real time. The information sets differ across models only if the models use different variables. Forecasts may also differ due to different estimation methods and different modeling assumptions. While the information set for the three small models and the Bayesian VAR is comprised of three time series, the information set of the CEE-SW model contains seven time series and the information set of our variant of the FRB-EDO model contains eleven time series. The particular time series and the sources for the real-time data set are described in appendix A2.

We re-estimate the models quarter-by-quarter with every arrival of a new data vintage. Thus, the newly estimated model specification uses parameter estimates $\hat{\theta}_t$ that are based on the information set I_t which contains the most recent data vintage available in quarter t. Of course, data on real GDP, the components of GDP and the associated deflators becomes available with a time lag and is not part of the current quarter t information set. Thus, current quarter growth and inflation estimates

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

are obtained using $t-1$ observations of those variables. The current quarter estimate is typically referred to as a *nowcast*, which is the ”forecast” at a horizon of zero quarters. The model forecasts for horizons $h \in (0,1,2,3,4)$ are computed under the assumption that $E[u_{t+h}|I_t] = 0$. They are generated by iterating over the following equation:

$$E[y^{obs}_{t+h}|I_t] = y(\hat{\theta}_t) + \hat{\lambda}_t + g_y(\hat{\theta}_t)^h y_{t-1}, \tag{10}$$

where a hat on the structural parameters θ and the subscript t denotes that they are estimated on the basis of the information set at time t, I_t, which contains the most recent releases of economic aggregates through quarter $t-1$. Recall also that the reduced form solution matrices g_y are functions of these estimates and change over time as new data vintages become available.
In conclusion of this section, it is instructive to summarize the different steps needed to generate diverse model forecasts:

1. Model Setup: create a model file with the model equations and add measurement equations that link the model to observable time series.
2. Solution: solve the model and write it in state space form
3. Data update: update the data with the current data vintage
4. Prior: add a prior distribution of the model parameters if necessary
5. Estimation: estimate the structural parameters by maximizing the likelihood or the posterior kernel
6. Forecast: compute forecasts by iterating over the solution matrices setting the expected value of future shocks to zero
7. Repeat steps 3 to 6 quarter-by-quarter for the time-period of interest.
8. Repeat steps 1 to 7 for different models possibly extending the information set by additional variables as required by the respective model.

4 An Illustration: Forecasting the 2001 recession

Next, we illustrate the real-time forecasting process with an example focusing on the last recession in the United States. According to the NBER Business Cycle Dating Committee the peak in economic activity in March 2001 marks the beginning of the last recession. The trough of economic activity in this recession was dated to be November 2001.

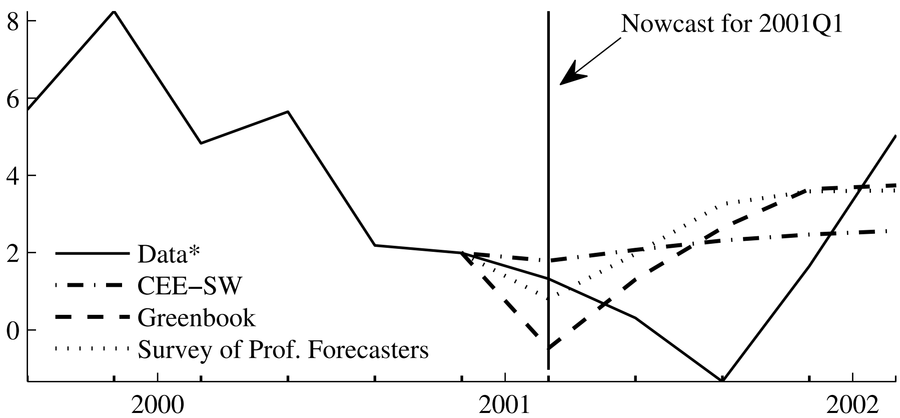

Figure 1: Real GDP Growth Forecast at the Start of the 2001 Recession (NBER defined peak: 2001Q1, NBER defined trough: 2001Q4). **Notes:***) The solid line shows data vintage 2001Q1 until 2000Q4 and revised data afterwards.

Figure 1 shows real output growth forecasts that were obtained on the basis of data available in the first quarter of 2001. The vertical solid line serves to indicate the current quarter. The nowcasts in 2001:Q1, of course, differs from the actual 2001:Q1 data that is released subsequently. The solid line in Figure 1 reports the actual data on annualized quarter on quarter output growth. This time series consists of the data vintage 2001:Q1 until the starting point of the nowcast/forecast in the fourth quarter 2000 and revised data from 2001:Q1 onwards. The revised GDP data is obtained from later data vintages.

GDP data is first released about one month after the end of the quarter to which the data refers, the so-called advance release. These data are then revised several times at the occasion of the preliminary release, final release, annual revisions and benchmark revisions. We follow Faust and Wright (2009) and use the data point in the vintage that was released two quarters after the quarter to which the data refer to as revised data. For example, revised data for 2001:Q1 is obtained by selecting the entry for 2001:Q1 from the data vintage released in 2001:Q3. Revised data for 2001:Q2 is obtained by using the entry for 2001:Q2 from the data vintage released in 2001:Q4, and so on. Hence, we do not attempt to forecast annual and benchmark revisions, because the models cannot predict changes in data definitions. The revised data against which we judge the accuracy of forecasts will typically correspond to the final NIPA release.

Three different forecasts are reported in Figure 1. The model-based forecast depicted by the dashed-dotted line is derived from the CEE-SW model. It is compared to the Fed's Greenbook forecast (dashed line) and the mean forecast from the Survey of Professional Forecasters (dotted line). All three forecasts imply a reduction in output growth in 2001:Q1, the current quarter, followed by a re-bound in subsequent quarters. The CEE-SW model only projects a slight decline in the growth rate compared to the larger declines implied by mean SPF forecast and the Greenbook. However, in this particular quarter the Greenbook nowcast of negative growth is far too pessimistic and ex-post the least accurate among the three nowcasts. As to the subseqent quarters, all three forecasts turned out to be mistaken in predicting an immediate re-bound starting in 2001:Q2. The economy deteriorated in the second and third quarter of 2001. The lowest quarterly output growth rate was reached in 2001Q3, and only afterwards the economy recovered.

Successive forecasts throughout the course of the 2001 recession are shown in Figure 2. The left-hand-side column of panels in Figure 2 compares the real-time forecasts generated with the CEE-SW model (solid line with square markers) to the Greenbook (dashed line) and SPF (dotted line) forecasts and the actual data (solid line). The top-left panel replicates Figure 1 with the 2001:Q1 forecasts. Moving down the columns the data vintages and forecasts are shifted forward quarter-by-quarter. The second left-hand-side panel indicates that the Greenbook and SPF nowcasts in 2001:Q2 were much closer to get the decline in GDP growth correctly than the CEE-SW model's nowcast. In 2001:Q3 the CEE-SW nowcast and forecasts for subsequent quarters are very similar to the Greenbook and SPF forecast. In 2001:Q4 the CEE-SW nowcast and forecasts clearly dominate the two expert forecasts in terms of accuracy. The Greenbook and mean SPF forecast implied a deepening of the recession. The revised data shows that instead a recovery took place as predicted by the model forecast. In 2002:Q1 the model nowcast is again more accurate. Also, the forecast for the third quarter is right on but at the expense of overshooting in the next two quarters.

The panels in the right-hand-side column of Figure 2 provide a comparison of the quarter-by-quarter forecasts generated from the six different macroeconomic models. The CEE-SW model is shown again along with the forecasts from the NK-DS, NK-WW, NK-Fu, BVAR-WW and FRB-EDO models. The solid line indicates again the data that is used as a benchmark for assessing the accuracy of the model forecasts. The model forecasts generally fail to forecast the downturn in the U.S. economy from the first to the third quarter of 2001. However, the mean SPF and Greenbook forecasts also largely miss the downturn. The model forecasts, however, perform relatively well with regard to the recovery, once the trough in 2001:Q3 has been reached. Model forecasts are quite heterogeneous with the extent of heterogeneity varying over time. Forecast differences narrow in 2001:Q2 and 2001:Q3 and widen again in 2001:Q4 and 2002:Q1.

5 Model-Based versus Expert Nowcasts and the 2008/09 Recession

The model-based forecasts shown in Figures 1 and 2 only used data vintages with the most recent entries concerning the quarter preceding the quarter in which the forecast is made. In this case, there are significant differences in the nowcasts of the models vis-a-vis the Greenbook and Survey of Professional Forecasters. In practice, however, there are many data series that are available on a

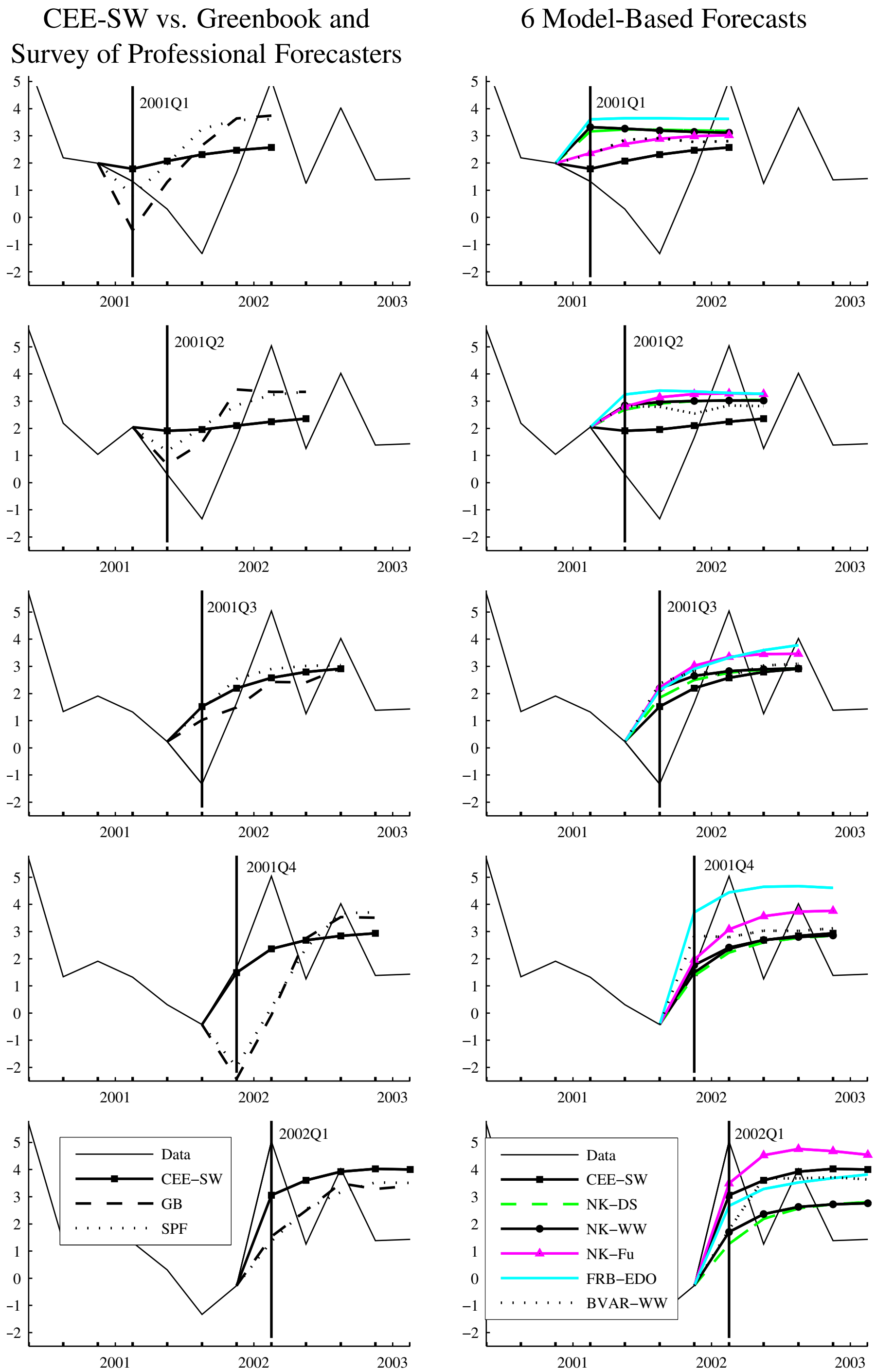

Figure 2: Real GDP Growth Forecasts for the 2001 Recession (NBER defined peak: 2001Q1, NBER defined trough: 2001Q4)

monthly, weekly or daily frequency that can be used to improve current-quarter estimates of GDP. Examples are industrial production, sales, unemployment, money, 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. In principle, there are technical methods available that allow the use of such data in combination with structural macroeconomic models. For example, Giannone, Monti, and Reichlin (2009) show how to incorporate such conjunctural analysis in structural models systematically. Employing such methods is beyond the scope of this paper. However, to approximate the effect of using more information in nowcasting we investigate the effect of using Greenbook or mean SPF nowcasts as a starting point for model-based forecasts regarding future quarters.

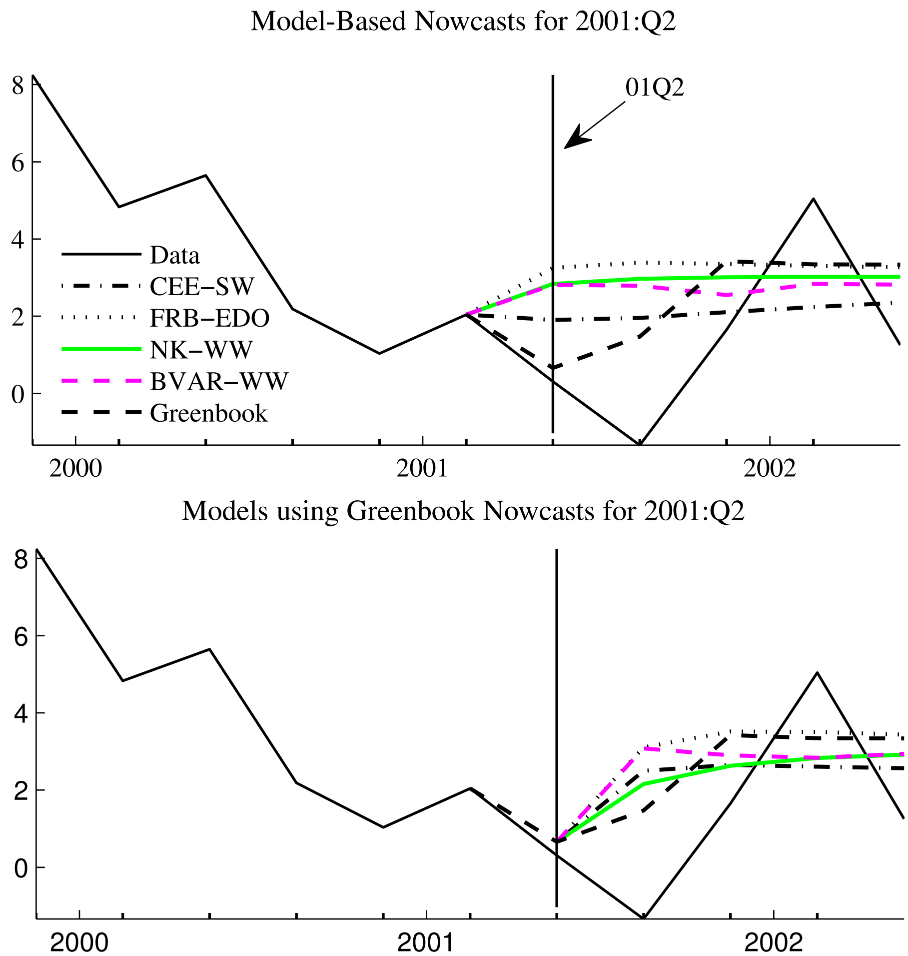

Figure 3: Real Output Growth forecast during the 2001 recession (NBER defined peak: 2001:Q1, NBER defined trough: 2001:Q4). **Notes:** In the upper panel the model-generated nowcast based on the information set with information on $t-1$ aggregates is used. In the lower panel the mean SPF nowcast forms the starting point for model-based forecasts regarding future quarters.

The difference between the use of model-based versus expert nowcasts as initial conditions for model-based forecasts for subsequent quarters is illustrated in Figure 3. The top panel in Figure 3 partly replicates the second right-hand-side panel in Figure 2. It shows the 2001:Q2 forecasts from the CEE-SW, FRB-EDO, NK-WW and BVAR-WW models in comparison to the Greenbook forecast (dashed line) and the revised data (solid line). As discussed previously, the Greenbook nowcasts in 2001:Q2 came much closer to capturing the start of the downturn than the model-based forecasts. Clearly, by that time it had become apparent to the Federal Reserve staff that the economy was deteriorating perhaps because of evidence obtained from higher-frequency data. The models miss this early evidence of the downturn as they are only using quarterly data concerning 2001:Q1.

The lower panel of Figure 3 displays the effect of using the Greenbook nowcast as the basis for the model-based forecasts. As a consequence, the model-based forecasts differ much less from each other than in the upper planel. The model-based forecasts one-quarter-ahead are more optimistic than the Greenbook forecast. The two quarter-ahead forecasts from the models, however, are somewhat below the Greenbook and a bit closer to the eventual realization of output growth.

Altogether, we investigate and compare successive forecasts throughout the five most recent recessions on the U.S. economy in this manner. Of course, at the current juncture it is of particular interest to investigate the accuracy and diversity of forecasts in the on-going recession. In 2008 and 2009 public criticism of economic forecasters for failing to predict the downturn that is now often referred to as ”The Great Recession” has been very pronounced.

Figure 4 provides a perspective on successive model-based forecasts relative to the mean SPF forecast (dash-dotted line) and the actual data (solid line) that has become available so far. The top row of panels shows forecasts made in the third quarter of 2008. Lower rows report subsequent forecasts quarter-by-quarter as new data vintages become available. In the panels of the left-hand-side column model-based nowcasts are generated from the most recent quarterly data vintage. In the right column, instead, mean SPF nowcasts are used to initialize the model forecasts.

As is apparent from the top left panel, professional forecasters, on average, failed to foresee the downturn as late as in the third quarter of 2008. The mean SPF forecast indicates a slowdown in the fourth quarter followed by a return to higher growth in the first quarter of 2009. Not surprisingly, this misdiagnosis has generated much public criticism. The model-based forecasts we generate based on the data vintage of 2008:Q3 would not have performed any better. In fact, they do not indicate any impending decline in economic activity. In the fourth quarter of 2008, however, the mean SPF nowcast and the model-based nowcast diverge dramatically. Following the Lehman debacle professional forecasters drastically revised their assessments downwards, and continued to do so in the first quarter of 2009.

Interestingly, from 2009:Q2 onwards the model-based forecasts perform quite well in predicting the

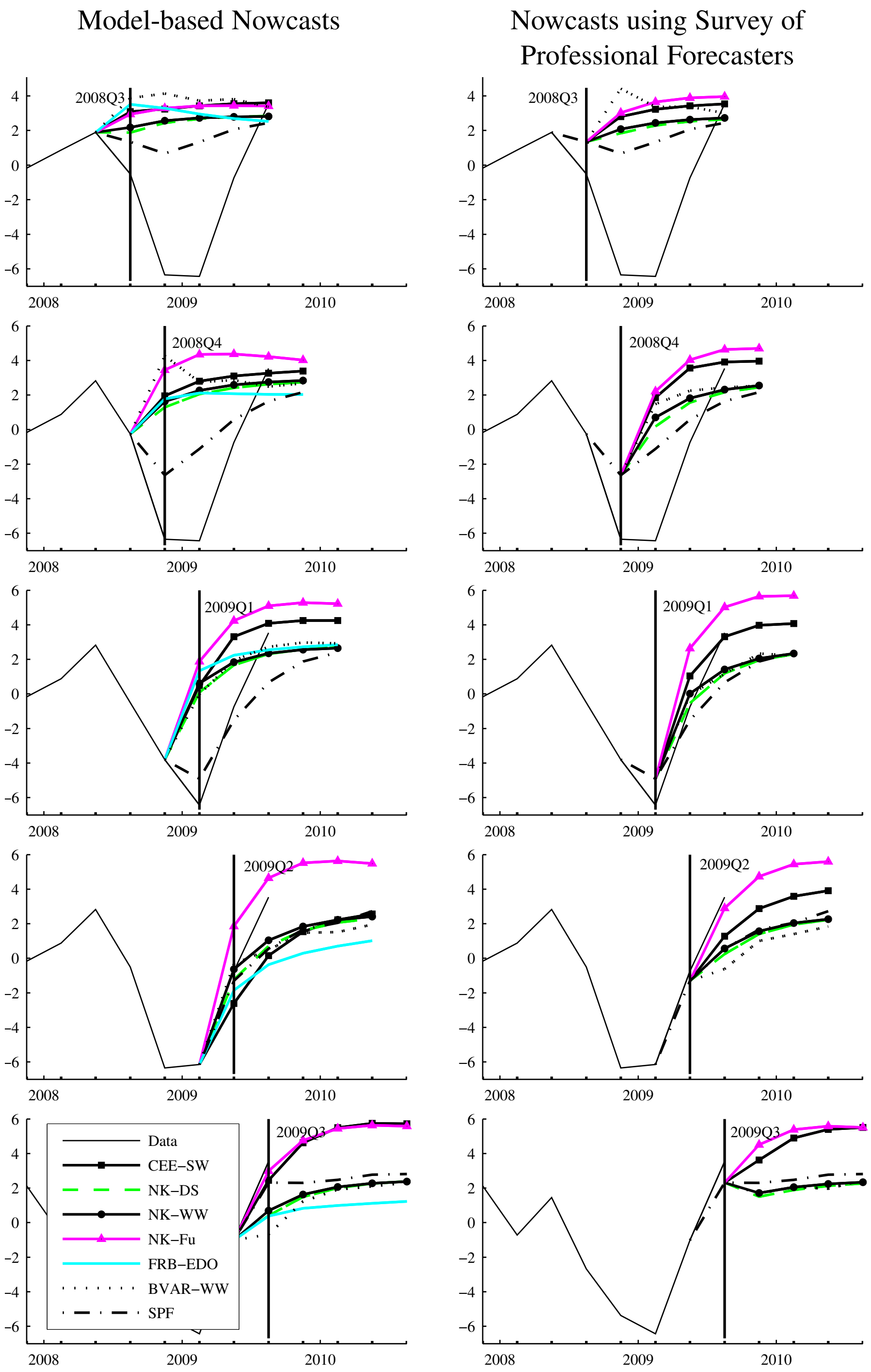

Figure 4: Real Output Growth forecast during the 2007-2009 recession (NBER defined peak: 2007Q4). **Notes:** In the left-hand-side panels the model-generated nowcast based on the information set with information on $t-1$ aggregates is used. In the right-hand-side panels the mean SPF nowcast forms the starting point for model-based forecasts regarding future quarters.

recovery of the U.S. economy. From that point onwards, several of the models deliver predictions that are very similar to the mean SPF forecast and match up with the subsequent data releases surprisingly well. An inspection of the right-hand-side panels suggests that initializing the model forecasts with the mean SPF nowcasts further strenghtens the models performance during the recovery phase. In this case, the 2009:Q1 forecast for the second and third quarter of 2009 that is implied by the CEE-SW, NK-WW and FRB-EDO models already looks surprisingly accurate relative to the data releases that have become available so far.

6 The Relative Accuracy of Model-Based and Expert Forecasts

For a systematic evaluation of forecast accuracy we compute the root mean squared errors (RMSE) of the nowcast and forecasts from one to four-quarters-ahead for each model during the five recessions. Each recession sample typically covers the period from 4 quarters prior to the trough determined by the NBER Business Cycle Dating Committee to 4 quarters after the trough.[4] The accuracy of the individual model forecasts is compared to the mean model forecast, that is the average of the six models, the mean SPF forecast and the Greenbook forecast. The RMSE for model m at forecasting horizon h given a recession sample that starts in period p and ends in period q is given by:

$$RMSE_m^h = \sqrt{\sum_{t=p}^{q} \left(E[y_{t+h}^{obs}|I_t^m] - y_{t+h}^{obs}\right)^2 / (q-p+1)}, \tag{11}$$

where I_t^m denotes the information set of a specific model m at time t. I_t^m includes the model equations and the data vintage for period t. y_{t+h}^{obs} denotes the data realizations h periods ahead.

Our findings are reported in Table 3. In most cases the model forecasts are on average less accurate than the Greenbook and mean SPF forecasts. Sometimes the best forecast is given by the Greenbook but at other times by the mean SPF forecast. The difference between the RMSEs of model and expert forecasts decreases with the forecast horizon. Structural models are therefore suitable for medium-term forecasts while expert forecasts incorporate additional information that helps improve nowcasts and near-term forecasts. An exception is the 2001 recession during which the quality of all forecasts is very similar. Root mean squared errors are lower during the 1990-91 recession and the 2001 recession than during the other recessions.

Among the structural models there is none that consistently outperforms the others. During a specific recession, the best forecasts at different horizons may also come from different models. Nevertheless, a comparison reveals some systematic differences. For example, the CEE-SW model delivers relatively good forecasts during the two recessions in the early 1980s and the 2001 recession. The BVAR-WW model exhibits a fairly high forecast accuracy in the 1990-91 recession and the NK-DS

[4]Exceptions are the 1980 and 2008/9 recessions. In first case, we start only 2 quarters prior to the trough because of data availability. In the second case, the trough is not yet determined. We start in 2007Q4 (peak) and end in 2009Q3.

model exihibts the lowest root-mean-squared errors during the on-going recession. The mean model forecast shown in the seventh column which averages the six model forecasts performs quite well. Most of the time it turns out to be fairly close to the best individual model forecast in terms of root mean squared error.

In addition, we have investigated the accuracy of inflation forecasts. Table 4 reports the associated root mean squared errors of nowcasts and forecasts for the five recession episodes. Again, the root-mean-squared errors at horizons from zero to four quarters into the future are recorded separately. The Federal Reserve's Greenbook forecast for inflation is almost always more accurate than the other forecasts including the mean forecast from the Survey of Professional Forecasters. Perhaps, the better performance of the Greenbook forecast reflects an informational advantage regarding the inflationary consequences of Federal Reserve policies and future policy intentions.

Interestingly, the quality of the mean model forecast of inflation is quite similar to the mean SPF forecast. The NK-WW model performs very well in the 1980-81 recession, the FRB-EDO model offers the best model-based inflation forecast in the 1982 recession, the BVAR-WW model in the

Table 3: RMSEs of Output Growth Forecasts

Sample / Horizon	NK-DS	NK-WW	CEE-SW	FRB-EDO	NK-Fu	BVAR-WW	Mean	GB	SPF
1980:1 - 1981:3									
0	7.19	7.12	6.42	5.64	6.88	6.46	5.13	5.05	–
1	7.28	7.20	5.59	5.95	6.78	7.63	5.59	6.65	–
2	5.56	5.67	5.24	5.77	7.43	8.69	5.70	5.54	–
3	5.50	5.67	4.33	4.92	5.62	6.28	4.56	6.11	–
4	5.43	5.57	4.45	4.39	5.56	7.33	4.84	5.32	–
1981:4 - 1983:4									
0	5.54	5.68	2.89	3.23	3.69	3.68	3.68	2.42	2.14
1	5.14	5.25	3.69	4.32	3.96	3.98	4.02	3.58	3.88
2	4.09	4.16	4.06	4.59	4.84	5.72	4.31	3.93	4.11
3	4.16	4.22	4.15	4.53	5.10	5.74	4.45	3.91	4.41
4	4.09	4.12	4.02	4.56	4.66	5.74	4.33	3.84	4.02
1990:1 - 1992:1									
0	2.82	3.01	3.22	1.80	2.92	1.76	2.50	1.27	1.12
1	3.15	3.22	3.94	2.06	3.79	2.24	2.98	2.09	1.45
2	3.08	3.13	4.00	2.15	3.84	2.38	2.99	2.34	2.06
3	3.13	3.14	3.90	2.38	3.81	2.56	3.03	2.31	2.54
4	2.79	2.78	3.56	2.30	3.73	2.32	2.80	2.18	2.37
2000:4 - 2002:4									
0	2.32	2.33	1.94	2.43	2.30	2.63	2.22	2.28	2.22
1	2.22	2.24	2.19	2.49	2.64	2.28	2.25	2.20	2.30
2	2.23	2.21	2.29	2.61	2.54	2.35	2.29	2.34	2.21
3	2.69	2.67	2.74	2.82	2.74	2.71	2.67	2.76	2.65
4	2.24	2.25	2.08	2.58	2.17	2.12	2.19	2.18	2.13
2007:4 - 2009:3									
0	3.58	3.75	3.78	4.05	4.37	4.42	3.91	–	1.94
1	4.36	4.43	4.81	4.72	5.18	4.95	4.69	–	3.30
2	4.78	4.83	4.89	4.85	5.36	5.05	4.94	–	4.11
3	5.20	5.21	5.35	5.13	5.66	5.29	5.29	–	4.80
4	5.56	5.55	5.85	5.29	5.91	5.61	5.62	–	5.39

1990-91 and 2001 recession and the CEE-SW model in the ongoing "Great Recession". The mean model forecast of inflation comes quite close to the best individual model forecast most of the time. As discussed in the preceding section, the quality of a forecast for the future also very much depends on how accurate the assessment of the current state of the economy is that forms the starting point for the forecast. The model forecasts lack information on specific events that have happened in the current quarter such as the failure of Lehman in the fall of 2008 nor do they make use of higher-frequency data that becomes available during the quarter ahead of quarterly GDP releases. Expert forecasts may take into account both types of information. Therefore, we check if the superior forecast performance of the expert forecasts is due to the same informational advantage that induces better nowcasts. As in the preceding section, we simply use the Greenbook nowcast (and for the latest recession the mean SPF nowcast) as initial conditions for the model-based forecasts. On this basis, we re-estimate the models and compute forecasts for horizons of one to four quarters into the future. Tables 5 and 6 report the associated root mean squared errors of output growth and inflation forecasts for the different recession episodes.

The GDP growth forecast improve for most models and horizons when the expert nowcast is added to

Table 4: RMSEs of Inflation Forecasts

Sample / Horizon	NK-DS	NK-WW	CEE-SW	FRB-EDO	NK-Fu	BVAR-WW	Mean	GB	SPF
1980:1 - 1981:3									
0	1.77	1.76	2.05	2.64	2.04	2.67	1.90	1.67	1.52
1	1.92	1.90	2.52	3.55	2.76	2.18	2.19	1.25	1.81
2	1.59	1.38	2.05	2.57	2.20	1.75	1.45	1.66	1.92
3	2.89	2.32	2.36	3.34	2.96	3.88	2.53	1.77	2.23
4	3.07	2.29	2.51	3.79	2.83	3.97	2.58	2.21	2.56
1981:4 - 1983:4									
0	1.90	1.76	1.69	1.37	2.41	1.49	1.58	1.12	1.13
1	2.71	2.24	1.98	1.47	2.16	2.24	1.98	1.32	1.76
2	2.63	1.99	1.89	1.29	1.81	2.13	1.70	1.26	1.68
3	2.85	2.01	2.10	1.31	2.07	2.31	1.80	1.07	1.95
4	2.87	1.95	2.26	1.22	1.61	2.46	1.67	1.48	2.06
1990:1 - 1992:1									
0	1.21	1.16	1.07	1.21	1.80	1.05	1.15	0.73	1.09
1	1.76	1.64	1.29	1.20	2.03	1.16	1.43	0.84	0.98
2	1.69	1.76	1.35	1.33	1.15	1.07	1.25	0.95	1.01
3	1.30	1.76	1.53	0.91	0.81	0.95	1.01	1.06	1.19
4	1.69	1.87	1.71	1.39	1.65	1.37	1.40	1.02	1.19
2000:4 - 2002:4									
0	1.08	1.05	1.04	1.27	1.17	0.90	0.98	0.56	0.70
1	1.18	1.15	1.12	1.43	1.26	0.92	1.07	0.87	0.87
2	1.35	1.38	1.16	1.50	1.48	1.11	1.19	0.70	0.92
3	1.42	1.49	1.21	1.75	1.63	1.16	1.28	0.75	0.93
4	1.45	1.59	1.07	1.64	1.83	1.30	1.27	0.78	0.98
2007:4 - 2009:3									
0	2.06	1.96	1.69	2.19	1.61	1.58	1.69	–	1.11
1	1.53	1.51	1.14	1.83	1.52	1.21	1.23	–	1.03
2	1.56	1.54	1.23	1.95	1.61	1.31	1.31	–	1.10
3	1.86	1.82	1.36	1.77	1.99	1.60	1.61	–	1.24
4	1.60	1.74	1.38	1.64	1.78	1.48	1.40	–	1.40

the models' information sets. An exception is the recession of 1980, probably because the Greenbook nowcasts were not very good during this period. The mean model forecast now even outperforms the Greenbook forecast in the 1980 and 2001 recessions. The mean model forecast also compares well to the mean SPF forecast in the 1981-82 and 2001 recessions. The Greenbook forecasts still perform best in 1981-82 and 1990-91 recessions, while the mean SPF forecast still appears to be the most accurate in the ongoing recession, for which no Greenbook data and forecasts are publicly available. With regard to forecasts of inflation, the addition of the expert nowcast to the information set of the model does not improve model-based forecasts quite as much as in the case of GDP forecasts. Also, the Greenbook forecast performance tends to remain superior to the model forecasts. Thus, one might speculate that the Federal Reserve staffs advantage in forcasting inflation is driven either by modeling assumptions or information regarding its own objectives and future policies.

Table 5: RMSEs of Output Growth Forecasts Initialized with Expert Nowcasts

Sample / Horizon	NK-DS	NK-WW	CEE-SW	FRB-EDO	NK-Fu	BVAR-WW	Mean	GB	SPF
1980:1 - 1981:3									
0	5.05	5.05	5.05	5.05	5.05	5.05	5.05	5.05	–
1	8.14	8.13	6.33	6.06	7.18	6.69	5.83	6.65	–
2	6.34	6.36	4.80	5.60	6.48	6.48	4.83	5.54	–
3	5.50	5.74	5.20	5.37	6.49	7.74	5.20	6.11	–
4	5.56	5.75	4.23	4.24	4.12	5.50	4.05	5.32	–
1981:4 - 1983:4									
0	2.42	2.42	2.42	2.42	2.42	2.42	2.42	2.42	2.14
1	4.28	4.50	3.74	3.27	3.80	3.23	3.54	3.58	3.88
2	3.99	4.05	4.22	4.09	3.98	4.09	3.86	3.93	4.11
3	4.14	4.23	4.05	4.52	4.64	4.87	4.25	3.91	4.41
4	4.08	4.11	4.07	4.67	4.73	4.89	4.28	3.84	4.02
1990:1 - 1992:1									
0	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.27	1.12
1	2.64	2.87	3.22	1.70	3.11	2.00	2.47	2.09	1.45
2	2.95	3.04	3.80	1.92	3.68	2.28	2.82	2.34	2.06
3	3.08	3.13	3.78	2.42	3.67	2.55	2.94	2.31	2.54
4	2.71	2.76	3.65	2.16	3.48	2.29	2.69	2.18	2.37
2000:4 - 2002:4									
0	2.28	2.28	2.28	2.28	2.28	2.28	2.28	2.28	2.22
1	2.17	2.15	2.31	2.84	2.06	2.48	2.23	2.20	2.30
2	2.09	2.10	2.11	2.61	2.35	1.98	2.11	2.34	2.21
3	2.74	2.72	2.68	2.98	2.51	2.66	2.65	2.76	2.65
4	2.25	2.26	2.08	2.40	2.24	2.30	2.19	2.18	2.13
2007:4 - 2009:3									
0	1.94	1.94	1.94	–	1.94	1.94	1.94	–	1.94
1	3.74	3.90	4.24	–	4.54	4.85	4.21	–	3.30
2	4.52	4.62	4.94	–	5.48	5.10	4.89	–	4.11
3	5.05	5.11	5.39	–	5.83	5.27	5.32	–	4.80
4	5.50	5.52	5.86	–	6.07	5.57	5.70	–	5.39

7 The Heterogeneity of Model-Based and Expert Forecasts

The examples with model-based forecasts of output growth in the 2001 and 2008/09 recessions shown previously in Figures 1 to 4 indicate a substantial degree of heterogeneity that varies over time during and across different recession episodes. In this section, we aim to document the extent and dynamics of forecast heterogeneity somewhat more systematically. To quantify forecast heterogeneity we compute the standard deviation of the cross section of individual forecasts for each horizon at any point in time. This standard deviation is defined as follows:

$$\sigma_t = \sqrt{\sum_{m=1}^{M} \left(E[y^{obs}_{t+h}|I^m_t] - \frac{1}{M} \sum_{m=1}^{M} E[y^{obs}_{t+h}|I^m_t] \right)^2 / (M-1)}, \quad (12)$$

where I^m_t denotes the information set of a specific model m at time t and M denotes the number of models used to forecast.

As a benchmark for comparison, we compute the same measure of forecast diversity for the cross section of individual expert forecasts from the Survey of Professional Forecasters. We only take into

Table 6: RMSEs of Inflation Forecasts Initialized with Expert Nowcasts

Sample / Horizon	NK-DS	NK-WW	CEE-SW	FRB-EDO	NK-Fu	BVAR-WW	Mean	GB	SPF
1980:1 - 1981:3									
0	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.67	1.52
1	2.73	2.59	2.57	2.76	2.97	2.94	2.59	1.25	1.81
2	2.89	2.56	2.49	2.53	2.76	3.33	2.59	1.66	1.92
3	2.70	1.86	1.98	1.39	1.48	2.71	1.73	1.77	2.23
4	4.02	2.92	2.54	3.00	3.15	4.94	3.22	2.21	2.56
1981:4 - 1983:4									
0	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.12	1.13
1	2.31	2.06	1.97	1.72	2.15	1.71	1.86	1.32	1.76
2	2.53	2.05	2.04	1.58	2.46	1.61	1.92	1.26	1.68
3	2.53	1.91	2.02	1.16	2.32	1.67	1.79	1.07	1.95
4	2.78	2.01	2.25	1.41	2.36	1.66	1.87	1.48	2.06
1990:1 - 1992:1									
0	0.73	0.73	0.73	0.73	0.73	0.73	0.73	0.73	1.09
1	1.03	1.10	1.01	0.94	1.77	0.93	1.03	0.84	0.98
2	1.42	1.58	1.36	0.81	1.61	1.04	1.23	0.95	1.01
3	1.49	1.77	1.63	1.11	0.89	0.93	1.20	1.06	1.19
4	1.31	1.70	1.62	1.34	0.87	1.07	1.16	1.02	1.19
2000:4 - 2002:4									
0	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.70
1	0.92	0.95	0.90	0.97	1.13	0.76	0.85	0.87	0.87
2	1.33	1.38	1.18	1.24	1.61	1.04	1.23	0.70	0.92
3	1.29	1.41	1.18	1.48	1.68	1.02	1.25	0.75	0.93
4	1.53	1.65	1.17	1.68	2.02	1.35	1.45	0.78	0.98
2007:4 - 2009:3									
0	1.11	1.11	1.11	–	1.11	1.11	1.11	–	1.11
1	1.15	1.19	1.00	–	1.48	1.11	1.10	–	1.03
2	1.28	1.37	1.17	–	1.56	1.22	1.28	–	1.10
3	1.50	1.61	1.30	–	1.87	1.49	1.51	–	1.24
4	1.69	1.81	1.39	–	1.92	1.59	1.65	–	1.40

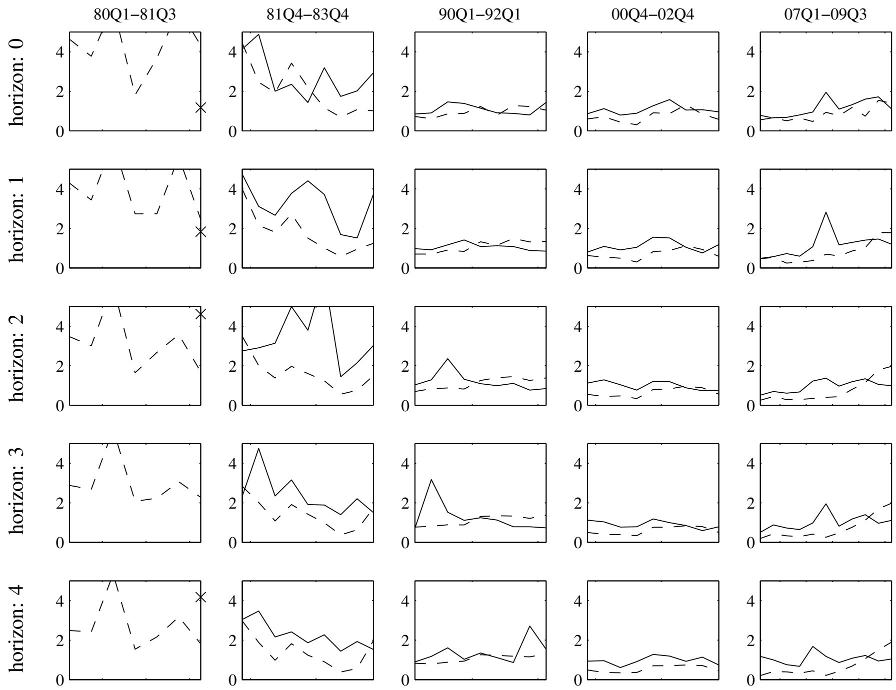

Figure 5: Standard Deviations of Output Growth Forecasts: Experts (solid) and Models (dashed)

account forecasters who contributed at least four forecasts during one of the recessions. As a result of this selection, the number of individual forecasts taken from the SPF ranges from 9 to over 50, compared to the 6 individual model forecasts.

Figures 5 and 6 display the standard deviations of model-based forecasts (dashed line) and professional forecasts (solid line). The rows show the different forecast horizons and the columns the different recessions. The dashed line shows the diversity of model forecasts while the solid line shows the diversity of survey forecasts. Output growth forecasts of the SPF start in 1981Q3 which is marked with an x.

The extent of heterogeneity of GDP growth and inflation forecasts is roughly in the same range for model-based and expert forecasts, although it is somewhat lower for the models relative to the experts. The latter finding might be attributed to the much smaller number of individual model forecasts. The diversity of forecasts among the six structural models provides an indication of the extent of disagreement that should be expected among practitioners that use state-of-the art models and methods. The disagreements between model-based forecasts can be traced to differences in particular modeling assumptions and to the range of variables considered.

We also conducted some robustness checks to find out whether the heterogeneity measured by the standard deviation is strongly influenced by outliers. We computed the range between the 0.166 and

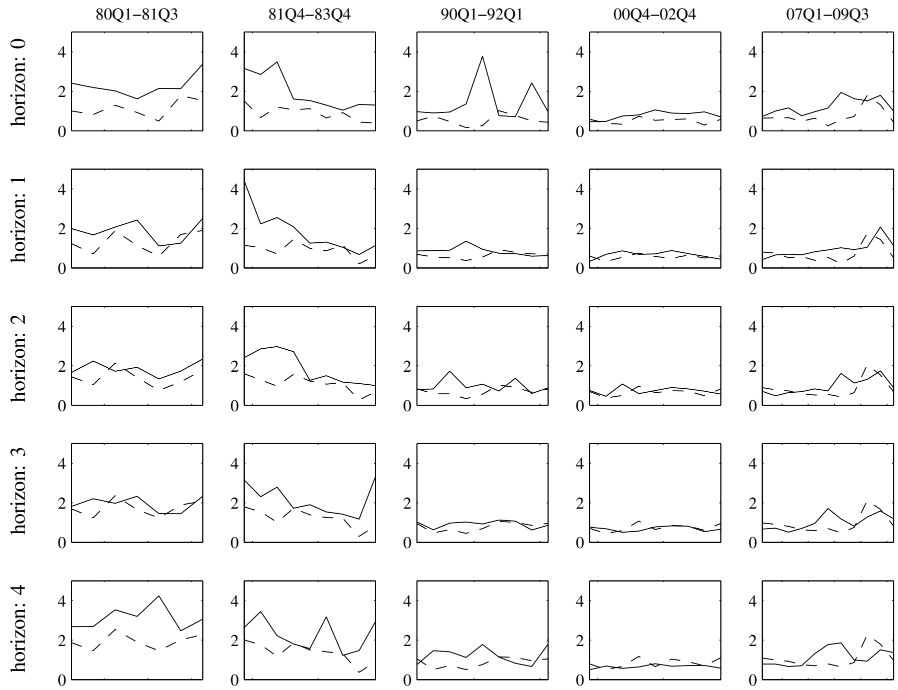

Figure 6: Standard Deviations of Inflation Forecasts: Survey (Black) and Models (Red)

0.833 quantile for model-based and professional forecasts, which means we drop the highest and the lowest model forecast, compute the range between the second highest and second lowest forecast and compare to the same measure obtained from expert forecasts. The results confirm the finding that the models generate a similar degree of diversity as observed in the Survey of Professional Forecasters.

In addition, it is apparent from Figures 5 and 6 that the extent of forecast heterogeneity varies a lot. For example, diversity in output growth forecasts is most pronounced in the 1980s recessions and much smaller in the 1990-91 and 2001 recessions. It increases again in the 2008/09 recession. At some occasions model-based and survey forecasts of GDP growth exhibit similar dynamics. Examples are the decline in the diversity of three- to four-quarter ahead forecasts over the course of the 1981-82 recession (last two panels in the second column), or the increase in diversity in the middle of the 2000-2002 period (fourth column of panels). However, we also observe some spikes in disagreement among forecasters in the SPF that do not appear in the model-based forecasts. Examples are found in the GDP growth forecasts in 1990 and 2008.

Another interesting aspect of heterogeneity concerns the accuracy of forecasts from individual forecasters. Some forecasters perform consistently better than average while others tend to make greater errors on average. Thus, we also compare the accuracy range among expert forecasters to the range among individual model forecasts. To this end, we compute the root mean squared error of the fore-

Table 7: RMSE of Best, Worst, and Average Output Growth Forecaster from Survey and Models

	Horizons:	0	1	2	3	4
1980:1 - 1981:3						
min RMSE	Survey / Models	– / 5.64	– / 5.59	– / 5.24	– / 4.33	– / 4.39
max RMSE	Survey / Models	– / 7.19	– / 7.63	– / 8.69	– / 6.28	– / 7.33
average RMSE	Survey / Models	– / 6.62	– / 6.74	– / 6.39	– / 5.39	– / 5.46
1981:4 - 1983:4						
min RMSE	Survey / Models	1.15 / 2.89	2.37 / 3.69	1.40 / 4.06	2.30 / 4.15	2.26 / 4.02
max RMSE	Survey / Models	10.33 / 5.68	15.12 / 5.25	18.91 / 5.72	9.77 / 5.74	10.22 / 5.74
average RMSE	Survey / Models	3.30 / 4.12	4.95 / 4.39	4.93 / 4.58	4.73 / 4.65	4.28 / 4.53
1990:1 - 1992:1						
min RMSE	Survey / Models	0.69 / 1.76	0.63 / 2.06	0.86 / 2.15	0.97 / 2.38	0.08 / 2.30
max RMSE	Survey / Models	2.36 / 3.22	2.74 / 3.94	4.67 / 4.00	5.23 / 3.90	8.54 / 3.73
average RMSE	Survey / Models	1.54 / 2.59	1.69 / 3.07	1.88 / 3.09	1.88 / 3.15	2.01 / 2.91
2000:4 - 2002:4						
min RMSE	Survey / Models	1.34 / 1.94	0.82 / 2.19	1.33 / 2.21	1.76 / 2.67	0.94 / 2.08
max RMSE	Survey / Models	4.72 / 2.63	3.49 / 2.64	4.22 / 2.61	3.76 / 2.82	3.10 / 2.58
average RMSE	Survey / Models	2.38 / 2.33	2.44 / 2.34	2.37 / 2.37	2.73 / 2.73	2.22 / 2.24
2007:4 - 2009:4						
min RMSE	Survey / Models	1.06 / 3.58	0.56 / 4.36	0.46 / 4.78	0.68 / 5.13	1.36 / 5.29
max RMSE	Survey / Models	12.95 / 4.42	12.03 / 5.18	7.77 / 5.36	9.28 / 5.66	7.70 / 5.91
average RMSE	Survey / Models	5.62 / 3.99	4.60 / 4.74	2.78 / 4.96	4.84 / 5.31	4.98 / 5.63

casts made by individual participants in the SPF for the different recession samples.

Table 7 reports the worst, best and the average RMSE of the individual expert forecasters during the five recession episodes. We only take into account those forecasters who contribute at least four forecasts for one of the recessions. Otherwise a very low RMSE can be achieved by forecasting only during times of little volatility. The average RMSE for output growth forecasts of survey participants and the six models is in a similar range, with the 1990-91 recession being an exception. During this recession the model forecasts are on average of worse quality than the forecasts of survey participants. The range of forecast accuracies is much wider in the SPF than among the six models. The SPF has some extreme outliers. The worst RMSE is as high as 18.91 in the 1981-82 recession for a forecast

Table 8: Best, Worst, and Average Inflation Forecaster from Survey and Models

	Horizons:	0	1	2	3	4
1980:1 - 1981:3						
min RMSE	Survey / Models	0.35 / 1.76	1.12 / 1.90	0.60 / 1.38	0.30 / 2.32	1.84 / 2.29
max RMSE	Survey / Models	5.81 / 2.67	4.92 / 3.55	4.50 / 2.57	4.46 / 3.88	8.49 / 3.97
average RMSE	Survey / Models	1.90 / 2.15	2.19 / 2.47	2.16 / 1.92	2.71 / 2.96	3.36 / 3.08
1981:4 - 1983:4						
min RMSE	Survey / Models	0.70 / 1.37	0.58 / 1.47	0.82 / 1.29	1.38 / 1.31	0.82 / 1.22
max RMSE	Survey / Models	6.52 / 2.41	9.36 / 2.71	6.42 / 2.63	9.58 / 2.85	6.56 / 2.87
average RMSE	Survey / Models	1.94 / 1.77	2.38 / 2.13	2.41 / 1.96	2.67 / 2.11	2.73 / 2.06
1990:1 - 1992:1						
min RMSE	Survey / Models	0.63 / 1.05	0.51 / 1.16	0.50 / 1.07	0.41 / 0.81	0.38 / 1.37
max RMSE	Survey / Models	8.40 / 1.80	2.27 / 2.03	2.98 / 1.76	2.35 / 1.76	2.46 / 1.87
average RMSE	Survey / Models	1.63 / 1.25	1.19 / 1.52	1.25 / 1.39	1.30 / 1.21	1.35 / 1.61
2000:4 - 2002:4						
min RMSE	Survey / Models	0.36 / 0.90	0.21 / 0.92	0.44 / 1.11	0.41 / 1.16	0.31 / 1.07
max RMSE	Survey / Models	2.50 / 1.27	1.83 / 1.43	2.73 / 1.50	2.18 / 1.75	1.85 / 1.83
average RMSE	Survey / Models	0.92 / 1.08	1.00 / 1.18	1.07 / 1.33	1.03 / 1.44	1.08 / 1.48
2007:4 - 2009:4						
min RMSE	Survey / Models	0.77 / 1.58	0.42 / 1.14	0.75 / 1.23	0.56 / 1.36	0.55 / 1.38
max RMSE	Survey / Models	6.00 / 2.19	2.52 / 1.83	4.21 / 1.95	4.31 / 1.99	4.99 / 1.78
average RMSE	Survey / Models	1.63 / 1.85	1.23 / 1.46	1.43 / 1.53	1.46 / 1.73	1.61 / 1.60

horizon of two quarters. The highest model RMSE of 8.69 is generated by the BVAR-WW model in the 1980 recession for a forecast horizon of two quarters. With few exceptions the maximal RMSE is higher among survey participants than among the models and the minimal RMSE is lower among survey participants than among models. The lowest survey RMSE is as low as 0.08 for a four-quarter horizon 4 in the 1990-91 recession. The lowest RMSE among the models is the nowcast of output growth in the 1990's recession with 1.76 and is also produced by the BVAR-WW model.

Table 8 reports the same statistics for the inflation forecasts. The average RMSE from the survey participants is always close to the average RMSE from the models. The best survey forecaster always performs better than the best model forecast. The worst survey forecast is with only one exception worse than the worst model forecast. The best survey RMSE is achieved for the 2001 recession for forecasting horizon of one quarter with a RMSE of 0.21. The best model RMSEs are given by 0.81 for the 1990-91 recession at a horizon of three quarters produced by the NK-Fu model and by 0.82 for the 2001 recession nowcast produced by the FRB-EDO model. We checked whether including the Greenbook or Survey nowcast in the information set for model-based forecasts changes these statistics. The models' minimal, maximal, and average RMSE decrease a little bit.

8 Conclusion

We conclude by summarizing the main findings of our analysis and comparison of model-based and expert forecasts.

Model-based forecasts, in particular the mean model forecast, compare quite well to the Greenbook forecasts and the mean SPF forecasts, especially at a horizon of three to four quarters into the future. Typically, model-based forecasts exhibit greater errors, but not that much greater than the Greenbook or professional forecasts. This result is somewhat surprising given that the macroeconomic models only take into account a small number of economic variables and the structural models incorporate theoretical restrictions that are known to be essential for evaluations of the impact of alternative policies but often considered a hindrance for effective forecasting purposes.

Model-based and professional forecasts perform badly at turning points. Interestingly, the model-based forecasts can do quite well during the recovery phase, sometimes even better than the Greenbook or mean-professional forecasts.

Professional forecasts typically make use of extensive survey information and higher-frequency indicators that help improve the estimate of current GDP prior to the first GDP release from the Bureau of Economic Analysis. Thus, it is not surprising if professional forecasts detect recessions a little earlier than model forecasts. However, model forecasts can be based on a similar footing in the terms of current information. Such an approach is presented, for example, by Giannone, Monti and Reichlin (2009). To approximate the effect of efficient now-casting we also conduct our comparisons between model-based and professional forecasts by starting from the professional now-cast. As a result, the gap between the two types of forecasts is further narrowed.

We quantify the extent of heterogeneity by means of the standard deviation across individual and model-based forecasts for a given forecasting horizon. The six model-based forecasts exhibit a broadly similar extent of forecast heterogeneity as the Survey of Professional forecasts.

The degree of forecast heterogeneity can change substantially over time. The standard deviations of model and professional forecasts vary over the course of the particular recession episodes that we examine as well as between different episodes. At some occasions the dynamics of forecast diversity derived from the two types of forecasts are quite similar.

Finally, we also compare the forecast quality of different forecasters and models. In other words, we compare the best, worst and average forecaster among models and professionals. In this case, the spread is much greater among the professionals in the SPF than among the different models. Thus, some professional forecasters are consistently worse than the worst model, while some others perform consistently better than the best model. Thus, we conclude that the range of accuracy of individual model forecasts does not approach the range observed in the Survey of Professional Forecasters.

References

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

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

Branch W, Evans G (2009) Monetary policy with heterogeneous expectations, forthcoming Economic Theory

Branch W, McGough B (2009) Business cycle amplification with heterogeneous expectations, forthcoming Economic Theory

Brock W, Hommes C (1998) Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics and Control 22:1235–1274

Capistran C, Timmermann A (2009) Disagreement and biases in inflation expectations. Journal of Money, Credit, and Banking 41:365–396

Chiarella C, Dieci R, He XZ (2007) Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework. Journal of Economic Behavior and Organization 62:408–427

Christiano LJ, Eichenbaum M, Evans CL (2005) Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113(1):1–45

De Grauwe P (2009) Animal spirit and monetary policy, forthcoming Economic Theory Special Issue?

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

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

Edge RM, Kiley MT, Laforte JP (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 RM, Kiley MT, Laforte JP (2008) Natural rate measures in an estimated dsge model of the u.s. economy. Journal of Economic Dynamics and Control 32:2512–2535

Edge RM, Kiley MT, Laforte JP (2009) A comparison of forecast performance between federal reserve staff forecasts, simple reduced form models and a dsge model, working Paper

Faust J, Wright JH (2009) Comparing greenbook and reduced form forecasts using a large realtime dataset. Journal of Business and Economic Statistics forthcoming

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

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

Fuhrer JC, 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

Giannone D, Monti F, Reichlin L (2009) The Science and Practice of Monetary Policy Today, Springer Science, chap Incorporating Conjunctural Analysis in Structural Models, pp 41–57

Giordani P, Söderlind P (2003) Inflation forecast uncertainty. European Economic Review 47:1037–1059

Goodfriend M, King RG (1997) The new neoclassical synthesis and the role of monetary policy. In: Bernanke BS, Rotemberg JJ (eds) National Bureau of Economic Research Macroeconomics Annual 1997, MIT Press, Cambridge, MA

Hamilton JD (1994) Time Series Analysis. Princeton University Press, Princeton, NJ.

Kimball M (1995) The quantitative analytics of the basic monetarist model. Journal of Money, Credit and Banking 27(4):1241–1277

Kurz M (1994a) On rational belief equilibria. Economic Theory 4:859–876

Kurz M (1994b) On the structure and diversity of rational beliefs. Economic Theory 4:877–900

Kurz M (1996) Rational beliefs and endogenous uncertainty: an introduction. Economic Theory 8:383–397

Kurz M (ed) (1997a) Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs. Springer Series in Economic Theory, No. 6, Springer Verlag

Kurz M (1997b) Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs, Springer Series in Economic Theory, No. 6, Springer Verlag, chap 1. Endogenous Economic Fluctuations and Rational Beliefs: A General Perspective

Kurz M (2008) Rational diverse beliefs and economic volatility, north holland. Handbook in Financial Economics forthcoming

Kurz M (2009) Handbook of financial markets: dynamics and evolution, North Holland, chap 8. Rational Diverse Beliefs and Market Volatility

Kurz M, Jin H, Motolese M (2003) Knowledge, Information and Expectations in Modern Macroeconomics: Essays In Honor of Edmund S. Phelps, Princeton University Press: Princeton, N.J., chap 10: Endogenous Fluctuations and the Role of Monetary Policy, pp 188 –227

Kurz M, Jin H, Motolese M (2005) The role of expectations in economic fluctuations and the efficacy of monetary policy. Journal of Economic Dynamics & Control 29:20172065

Rotemberg JJ, 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

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

Walsh CE (2003) Monetary theory and policy. The MIT Press, Cambridge

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

Wieland V (ed) (2009) The Science and Practice of Monetary Policy Today. Springer Science

Woodford M (2003) Interest and prices: Foundations of a theory of monetary policy. Princeton University Press, Princeton

Appendix A1: The Macroeconomic Models Used to Generate Forecasts

This appendix provides a more detailed description of the six macroeconomic models that are used in this paper to generate real-time forecasts based on the data available at the time the particular forecast is made.

NK-DS Model: The first model is described in Del Negro and Schorfheide (2004). An expectational IS equation and a policy rule together determine aggregate demand, while a New-Keynesian Phillips curve determines aggregate supply. Persistence in the demand curve is generated by serially correlated government spending and technology shocks - the latter enters demand via consumption habit formation - and an interest rate smoothing term in the monetary policy rule. The Phillips curve is derived from monopolistically competitive firms that have quadratic menu costs. Calvo-style staggered price contracts would lead to the same equation. Wang (2009) shows that the small number of frictions is sufficient to provide reasonable output growth and inflation forecasts.

NK-WW model: The second New-Keynesian model generalizes the baseline simple New-Keynesian model. To allow for richer output and inflation dynamics we add serially correlated preference and mark-up shocks.

NK-Fu Model: The third model is described in Fuhrer (1997). 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 and is able to match the strong inflation persistence observed in U.S. data. The model also exhibits a high degree of inertia with respect to aggregate demand. The aggregate demand curve is a single backward looking equation corresponding to an IS curve. The demand, supply and monetary policy shock are not serially correlated. The three small models are estimated on time series for output growth, inflation and the federal funds rate.

Medium scale DSGE models: The two medium scale DSGE models extend the baseline New-Keynesian model by fully incorporating recent advances in terms of microeconomic foundations. They contain a large number of frictions and structural shocks. The models are described in Smets and Wouters (2007) and Edge et al (2008). They are closed economy models and build on the work by Christiano et al (2005). Physical capital is included in the production function and capital formation is endogenized. Labor supply is modelled explicitly. Nominal frictions include sticky prices and wages and inflation and wage indexation. Real frictions include consumption habit formation, investment adjustment costs and variable capital utilization.

CEE-SW Model: The version of the Christiano et al (2005) model estimated by Smets and Wouters (2007) features nonseparable utility and fixed costs in production. The Dixit-Stiglitz aggregator is replaced with the aggregator by Kimball (1995) which leads to a non-constant elasticity of demand. The model consists of 14 equations that include forward looking consumption, investment, price and wage setting equations as well as several identities. Including seven structural shocks makes it possible to fit the model to seven empirical time series: output growth, inflation, federal funds rate, consumption, investment, hours and wages. The shocks are a total factor productivity shock, a risk premium shock, an investment-specific technology shock, wage and price mark-up shocks, an exogenous government spending shock and a monetary policy shock. All shock processes are serially correlated.

FRB-EDO Model: The model by Edge et al (2008) is a more disaggregated model. It features two production sectors, which differ in their pace of technological progress. This structure can capture the different growth rates and relative prices observed in the data. Accordingly, the expenditure site 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. We estimate a variant of the FRB-EDO model that abstracts from the flexible price allocation, but is for the rest of the model as close to the documentation (Edge, Kiley, and Laforte, 2007) as possible. Our version is not able to replicate the figures in the documentation exactly, but is reasonably close. The model is able to capture different cyclical properties in these four expenditure categories. It includes 14 structural shocks: technology shocks, price and wage mark-up shocks, preference shocks, capital efficiency shocks, an external spending shock and a monetary policy shock. The model is estimated on eleven empirical 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.

BVAR-WW: In addition to the five structural models we estimate a VAR on output growth, inflation and the Federal Funds Rate using 4 lags. The VAR is a more general description of the data than the DSGE models as it imposes little restrictions on the data generating process. All variables are treated symmetrically and therefore the VAR incorporates no behavioral interpretations of parameters or equations. Unrestricted VARs are heavily overparametrized and therefore not suitable for forecasting. We therefore use a Minnesota prior (see Doan et al, 1984) to shrink the parameters towards zero.

Appendix A2: The Quarterly Vintage Database

This appendix describes the data series and the data sources for the quarterly data vintages that form the basis of the quarterly real-time re-estimation of macroeconomic models over the business cycle in this paper.

TO BE ADDED.