

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

May 20, 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, two anonymous referees and the participants at the Stanford Institute for Theoretical Economics Workshop on Diverse Beliefs and the MONFISPOL Networks Stresa workshop for very helpful comments on this paper. The usual disclaimer applies.

[†]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 expert forecasts¹ of GDP and inflation (see 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 et al. (2005), Chiarella et al. (2007), Branch and McGough (2011), Branch and Evans (2011) and De Grauwe (2011).

Forecast heterogeneity arises for several reasons. First of all, forecasters need a forecast-generating framework. Such a framework may be a fully developed economic structure, a non-structural collection of statistical relationships or a simple rule-of-thumb. The particular modeling assumptions embedded in this forecasting framework represent an important source of belief heterogeneity. Another source of heterogeneity is the information used by the forecaster. Information sets may differ in terms of the number of economic aggregates or prices for which the forecasters collect data and the timeliness of the data vintage. The data is needed to estimate the state of the economy and the parameters of the forecasting framework.

While expert forecasts are published in various surveys, the underlying modeling assumptions, information sets and parameter estimates are not publicly available. Instead, this paper uses six different macroeconomic models of the U.S. economy to generate output and inflation forecasts and investigate the impact of modeling assumptions, information sets and parameter estimates on forecast precision and heterogeneity.² The precision and diversity of expert forecasts from the Survey of Professional Forecasters (SPF) and the Federal Reserve's Greenbook are used as benchmark for comparison.³ 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 a role in these shifts in economic activity.

Among the six macroeconomic models considered in this paper are three small-scale New-Keynesian

¹Expert forecasts are available via surveys such as Bluechip Economic Indicators by Aspen Publishers or the Survey of Professional Forecasters by the Federal Reserve Bank at Philadelphia.

²We draw on a recent research initiative that aims to build a database of macroeconomic models and offers a new comparative approach to model building and the search for macroeconomic policies that are robust under model uncertainty (see Taylor and Wieland (2009) and Wieland et al. (2009)).

³The SPF is conducted quarterly and contains responses by 30 to 50 professional forecasters. It was initiated in 1968 by the American Statistical Association and the NBER and is administered by the FRB Philadelphia since 1990. The Greenbook is not a survey. It contains a single forecast produced by the staff of the Board of Governors of the Federal Reserve System in Washington DC and becomes publicly available with a five-year lag.

models that differ in terms of structural assumptions, a non-structural Bayesian VAR model, and two medium-scale New-Keynesian dynamic stochastic general equilibrium (DSGE) models of the type currently used by leading central banks. The four small models are estimated to fit three macroeconomic time series: real GDP growth, inflation measured by the GDP deflator and the federal funds rate. The two medium-scale models are estimated with data for 7 and 11 variables, respectively. These variables include consumption, investment, wages and hours worked. The largest model even accounts for the breakdown in durables versus non-durables and services consumption, residential versus business investment, and the related deflators. We consider each of the six macroeconomic models as a reasonable forecast-generator. Such models are used at central banks and similar models may also be used by professionals in the private sector. Although the five structural models all embody the popular modeling assumption of homogenous rational expectations, they can be used together to generate an estimate of forecast heterogeneity due to differences in other modeling assumptions, information sets and parameter estimates. The properties of these models are discussed in more detail in the next section.

To render model-based forecasts comparable to historical SPF and Greenbook 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 real-time data set that contains all the historical quarterly vintages of the 11 time series used in the largest model. Every quarter we re-estimate all the model parameters on the basis of the data vintage that was available at that exact point in time. Using this parameterization we compute an estimate of the current state of the economy—the so-called *nowcast*— and forecasts for one 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. This assessment is obtained for periods surrounding recessions of the U.S. economy in 2008/09, 2001, 1990/91, 1981/82 and 1980. Forecasts are generated starting 4 quarters prior to the trough determined by the NBER Business Cycle Dating Committee up to 4 quarters after the trough.⁴

The approach taken in this paper breaks new ground in several respects. First, to our knowledge there exists no comparable assessment of the forecasting accuracy of multiple structural macroeconomic models with historical data vintages. Real-time forecasts of non-structural time series models have been compared recently by Faust and Wright (2009) and in earlier work by Bernanke and Boivin (2003). Edge et al. (2010) have provided an assessment of the real-time forecasting performance of a single structural model. Furthermore, this paper is the first attempt to quantify the heterogeneity of model forecasts and compare them to survey forecasts in order to learn more about the extent,

⁴Exceptions are the 1980 and 2008/9 recessions. In the first case, we start only 2 quarters prior to the trough because some data is not available for earlier vintages. In the second case, the trough is not yet determined. We start in 2007Q4 and end in 2009Q3.

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 extent of forecast heterogeneity is similar for model and professional forecasts but varies substantially over time. This variation itself may constitute 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. Of course, the models used by professional forecasters may differ from our models. Furthermore, New-Keynesian DSGE models have only been developed in the last decade and would not have been available to forecasters in earlier recessions. However, non-structural VAR models such as the Bayesian VAR were already in use in the 1980s and the model of Fuhrer (1997) is a good example of the type of structural models with rational expectations that have been used since the early 1990s. Even if most private sector forecasters still favor traditional structural models over the New-Keynesian DSGE models with microeconomic foundations preferred by academia and central banks, the two types of models exhibit some similar reduced-form relationships such as price and wage-inflation Phillips curves and aggregate demand equations with a mixture of backward- and forward-looking components. Thus, our findings can be taken as an indication that much of the observed time variation in forecast heterogeneity may be explained by disagreement about appropriate modeling assumptions and differences in parameter estimates rather than irrationality of particular forecasters.

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 total, we consider six different models of the U.S. economy. One of the models is a simple vector autoregression model (VAR) that incorporates no behavioral interpretations of parameters or equa-

tions. The other five models are structural representations of the U.S. economy. Table 1 summarizes the most important model features, while appendix A1 provides a complete description of the model equations.

Table 1: Model Overview

Name/Reference	Short Name	Type	Observable Variables	Original Authors' Sample
Bayesian VAR estimated in this paper	BVAR-WW	Bayesian VAR with 4 lags and Minnesota priors	3: output growth, inflation, interest rate	
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
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	
Christiano et al. (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 the 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

The VAR model is estimated with four lags of output growth, inflation and the federal funds rate. It is well-known that unrestricted VARs are heavily over-parameterized and therefore not very useful for forecasting purposes. As proposed by Doan et al. (1984) we use a Bayesian approach with so-called Minnesota prior 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. The extension *WW* is meant to indicate that we have estimated this model without reference to an earlier parameterization by other authors. Nevertheless, such models have been used in forecasting by many practitioners at least since the early 1980s, that is throughout all the recessions studied in our forecast comparison.

The structural models we have chosen reflect the developments in macroeconomic modeling in the last two decades. The model of Fuhrer (1997) is a good example of the New-Keynesian models that were developed in the 1980s and early 1990s.⁵ While academics still focused primarily on developing

⁵These models combined rational expectations and nominal rigidities as in the seminal paper of Taylor (1979). For other

the microeconomic foundations of real business cycle theory, these models became quite popular among central bank researchers and practitioners. They took into account adaptive and forward-looking behavior of market participants, real effects of monetary policy and output and inflation persistence. Fuhrer (1997) used maximum likelihood estimation to parameterize the model and we follow the same approach in re-estimating this model in the present paper. It is referred to as the *NK-Fu model* in our analysis.

The New-Keynesian model laid out by Rotemberg and Woodford (1997) and Goodfriend and King (1997) and developed in detail in Woodford (2003) and Walsh (2003) accounts more systematically for microeconomic foundations in terms of the optimizing and forward-looking behavior of representative households and firms. Such a framework is particularly useful for quantifying likely market responses to changes in macroeconomic policies as emphasized in the famous Lucas critique. The New-Keynesian model also incorporates restrictions in terms of monopolistic competition and price rigidity that imply important interactions between nominal and real economic variables. It has quickly become the principal workhorse model of monetary economics in the last decade.⁶ Key model variables are output, inflation and interest rates just as in the *BVAR-WW* and *NK-Fu* models, but the microeconomic foundations imply additional restrictions on the reduced-form VAR representation of this model. We consider two empirical implementations. The first specification is taken from Del Negro and Schorfheide (2004). They 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 in terms of the modeling assumptions regarding the particular economic shocks that are the source of fluctuations. It is also estimated with Bayesian methods and termed the *NK-WW model*.

Christiano, Eichenbaum and Evans (2005) extended 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 this end, they introduced 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. Smets and Wouters (2003, 2007) extended and estimated the model of Christiano, Eichenbaum and Evans with Bayesian methods to fit key macroeconomic series. We generate forecasts from a version of this model estimated with Bayesian methods and refer to it as the

examples see the model comparison projects of Bryant et al. (1988), Bryant et al. (1989), Klein (1991), and Bryant et al. (1993). Taylor (1993) already presented an estimated multi-country model of the G-7 economies of this type.

⁶For recent discussions of the application of the New-Keynesian approach in practical monetary policy see Wieland (2009).

CEE-SW model in the following. DSGE modeling has rapidly gained in popularity and many central banks have 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 et al. (2008). Following these authors we refer to it as the *FRB-EDO model*.

The two medium-scale models are fit to 7 and 11 economic time series, respectively. The CEE-SW model is estimated with data 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 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 data used in estimation covers 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.

3 Forecasting Methodology

This section demonstrates how the forecasts are computed. Three aspects are best distinguished and discussed separately: model specification and solution, parameter estimation, and the sequence of steps necessary to generate quarter-by-quarter forecasts.

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

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

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(x_t - g_t) \quad (2)$$

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

The notation of equations, variables and parameters is the same as in Del Negro and Schorfheide (2004). Variables are defined as percentage deviations from their steady state level. 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 (not shown). The monetary policy shock $\varepsilon_{R,t}$ is iid-normally

distributed. $(\beta, \tau, \gamma, r^*, \pi^*, \kappa, \rho_R, \psi_1, \psi_2)$ represent model parameters that need to be estimated. The vector of parameters also includes the AR parameters (ρ_g, ρ_z) governing the dynamics of economic shocks and the standard deviations of the associated innovations, $(\sigma_R, \sigma_g, \sigma_z)$.

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:

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

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

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

YGR_t denotes the first difference of the log of GDP, $INFL_t$ the first difference of the log GDP deflator, and INT_t the quarterly federal funds rate. 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:

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

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

$$E(u_t u_t') = Q(\theta), \quad (9)$$

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. These include the shock variances, so that Q also depends on elements of θ . A state space representation of this form is derived for each forecasting model and the notation in equations (7), (8) and (9) is general enough to apply to all the structural models considered. As an example, Table 2 shows how to link the variables and parameters in the state space representation to those in the Del Negro & Schorfheide model.

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 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 that is driven by labor-augmenting technological progress no separate de-trending of the data is necessary.

Table 2: State Space Representation and Model Equations

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

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 the reduced-form coefficients of the state-space representations are nonlinear functions of 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).⁷ 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 $\ln \mathcal{L}(\theta|Y^T) + \ln p(\theta)$ 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 robustness check we compared point forecasts obtained from the posterior mean and posterior mode in several cases. To this end, we simulated the posterior distribution using the Metropolis-Hastings-Algorithm. Since the two alternative point forecasts were quite similar we relied on the posterior mode for forecast generation in the remainder of our analysis so as to keep the computational burden resulting from the large number of model re-estimations manageable.

In estimating the Bayesian VAR we follow Doan et al. (1984) and use the so-called Minnesota prior to avoid over-parameterization. This prior implies shrinking the parameters towards zero by assuming that the price level, real output and the interest rate follow independent random walks.

⁷We consider only unique stable solutions. If the Blanchard-Kahn conditions are violated we set the likelihood equal to zero.

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 become available with a time lag and is not part of the current quarter t information set. Current quarter estimates of economic growth and inflation are obtained using $t - 1$ observations of those variables. The current quarter estimate is typically referred to as the *nowcast*, that 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 expected future shocks are equal to zero, $E[u_{t+h}|I_t] = 0$. They are generated by iterating over the following equation:

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

A hat on the structural parameters θ and the subscript t denote 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.

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 while extending the information set with 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 2001 recession in the United States. According to the NBER Business Cycle Dating Committee a peak in economic activity in March 2001 was followed by a trough in November 2001.

Figure 1 shows real output growth forecasts that were obtained on the basis of data available in the first quarter of 2001. The vertical line serves to indicate the current quarter. The nowcasts in 2001:Q1, of course, differ 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 of 2000 and revised data from 2001:Q1 onwards. The revised GDP data is drawn from later data vintages.

GDP data is first released about one month after the end of the quarter to which the data refers, the

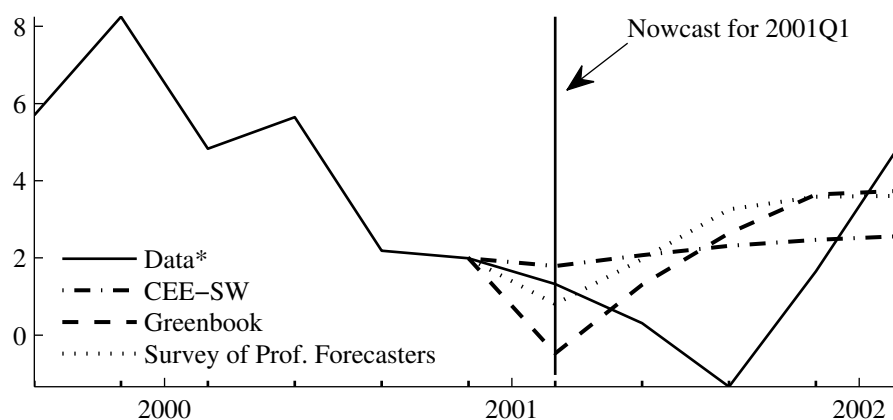


Figure 1: Real GDP Growth Forecast at the Start of the 2001 Recession (NBER defined peak: 2001Q1, NBER defined trough: 2001Q4).

Notes: *) The solid line shows data vintage 2001Q1 until 2000Q4 and revised data afterwards.

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 SPF forecast (dotted line). The SPF is a quarterly survey of professional macroeconomic forecasters conducted by the Federal Reserve Bank of Philadelphia. Typically, 30 to 50 respondents report projections of several key macroeconomic variables.⁸ Since these experts tend to earn their living in the forecasting business and may be expected to put serious effort in the production of the forecast, we consider it a reasonable benchmark for comparison with our model forecasts. Of course, it is well known that there exist incentives not to report the best possible forecast in such a survey.⁹ For this reason, we also consider the Greenbook forecast prepared by the staff of the Board of Governors of the Federal Reserve System for the Federal Open Market Committee.¹⁰

All three forecasts imply a reduction in output growth in 2001:Q1, the current quarter, followed by a re-bounce 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 the least accurate among the three nowcasts. As to the subsequent quarters, all three forecasts turned out to be mistaken in predicting an immediate re-bounce starting in 2001:Q2. The economy deteriorated in the

⁸Other surveys include Bluechip Economic Indicators, the Michigan Survey of Consumer Attitudes and Behavior and the Livingston Survey. Livingston and Bluechip are surveys of professionals like the SPF. Bluechip is not available free of charge. The Livingston survey is only conducted semi-annually. The Michigan survey reports assessments of 1000 to 3000 households. Mankiw et al. (2004) compare inflation expectations from these different surveys: median inflation expectations are relatively accurate and similar for the different surveys. Histograms show substantial disagreement; especially among consumers. There are extreme outliers that show up in long tails of the forecast distribution. Disagreement varies dramatically over time but similarly for consumers and professionals. Mishkin (2004) is sceptical of household surveys and notes that households have no incentive to compute detailed forecasts to answer survey questions about their expectations. Given the long tail in forecast distributions, he questions whether respondents with extreme expectations behave in a way consistent with these expectations. Professional forecasters, who make their living in this business, will put serious effort in computing a good forecast.

⁹Forecasters have incentives to publish a forecast close to the consensus (Scharfstein and Stein, 1990; Lamont, 2002) as well as to publish a distinct forecast (Laster et al., 1999).

¹⁰Greenbook projections are prepared by the Federal Reserve's staff before each FOMC meeting and have been found to dominate forecasts from other professional forecasters in terms of forecasting accuracy (Romer and Romer, 2000; Sims, 2002; Bernanke and Boivin, 2003). They are made public with a five-year lag.

second and third quarter of 2001. The lowest quarterly output growth rate was reached in 2001:Q3, after which 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 the actual decline in GDP growth 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. At that point 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 target although 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 forecast is shown together with the forecasts from the NK-DS, NK-WW, NK-Fu, BVAR-WW and FRB-EDO models. The solid line again indicates the data that is used as 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 use quarterly data vintages where the most recent data entries concern the quarter preceding the quarter in which the forecast is made. 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 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. The use of higher-frequency data may well be the main reason for better nowcasts by the Greenbook and Survey of Professional Forecasters compared to our six models.

CEE-SW vs. Greenbook and Survey of Professional Forecasters

6 Model-Based Forecasts

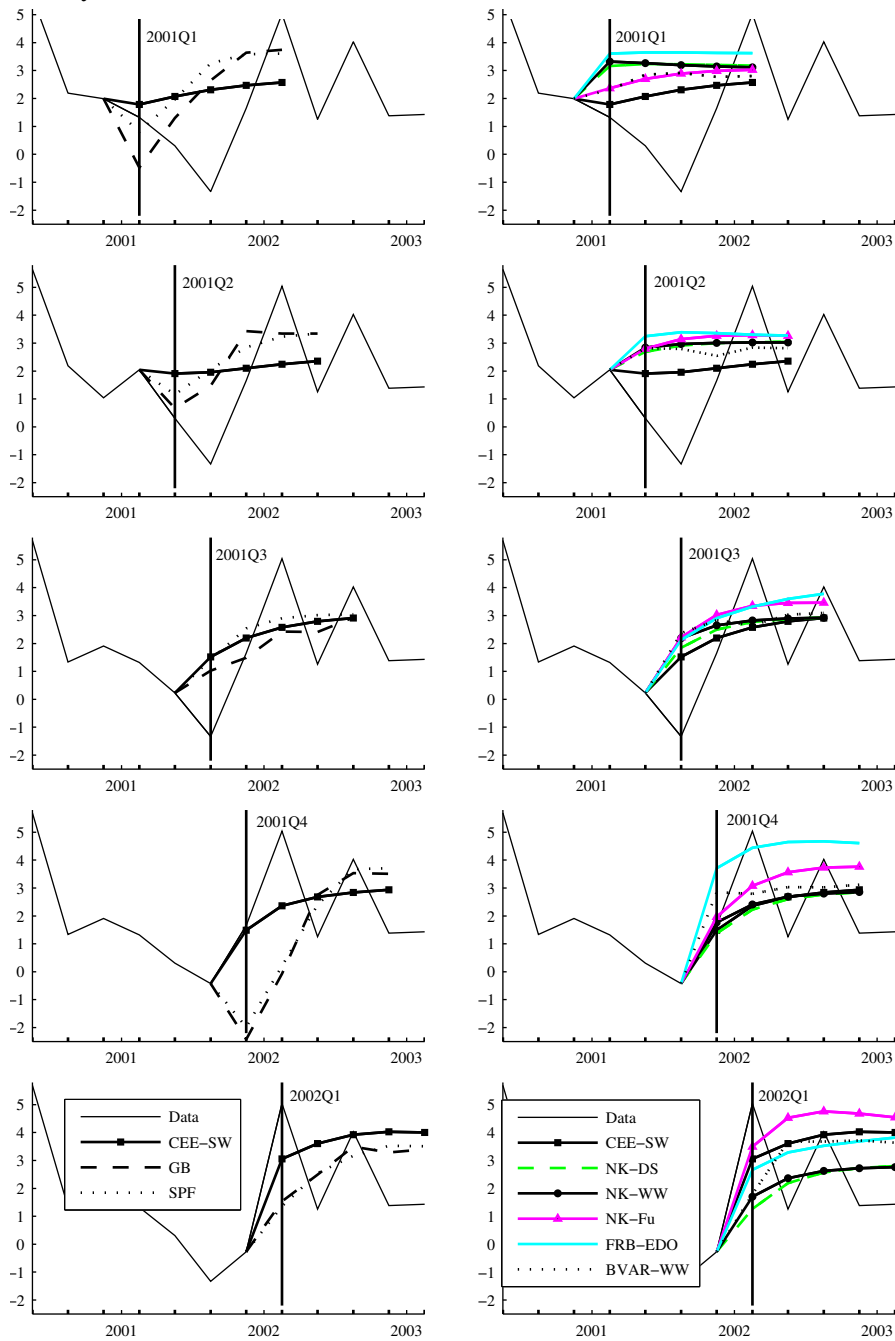


Figure 2: Real GDP Growth Forecasts for the 2001 Recession (NBER defined peak: 2001Q1, NBER defined trough: 2001Q4)

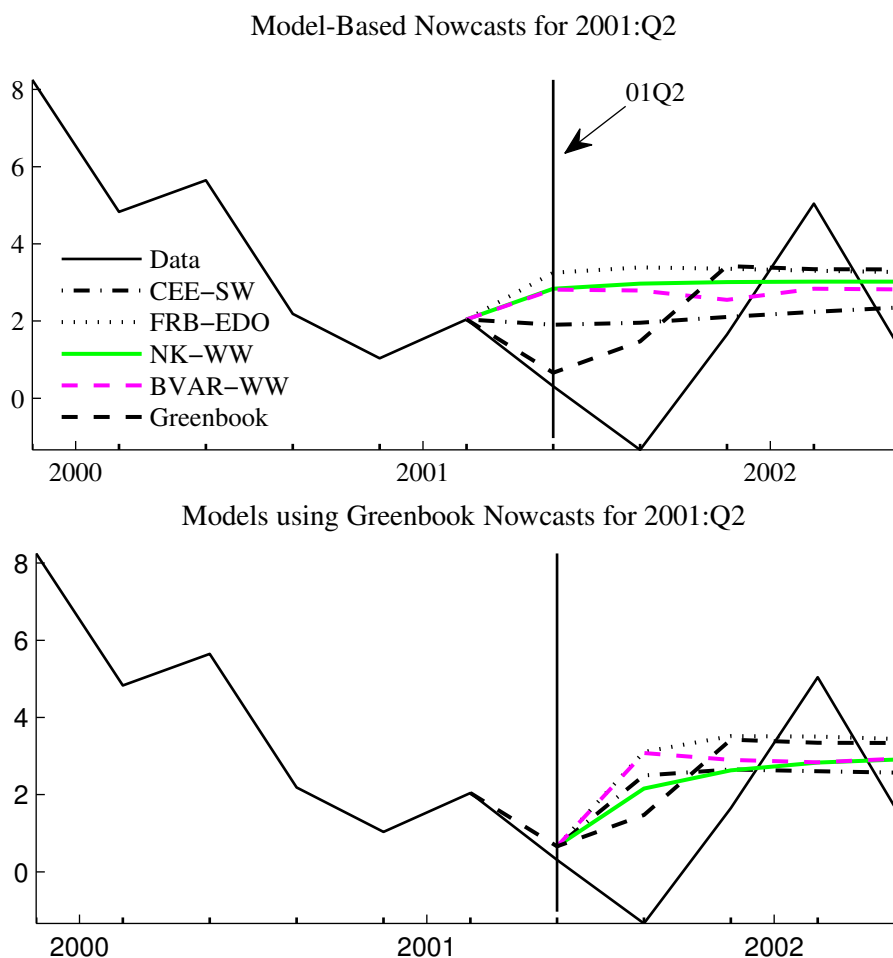


Figure 3: Real Output Growth forecast during the 2001 recession (NBER defined peak: 2001:Q1, NBER defined trough: 2001:Q4).

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 Greenbook nowcast forms the starting point for model-based forecasts regarding future quarters.

In principle, there exist methods for using higher frequency data in combination with quarterly structural macroeconomic models. For example, Giannone et al. (2009) show how to incorporate such conjunctural analysis systematically in structural models. Employing such methods, however, is beyond the scope of this paper. Instead, we approximate the use of higher-frequency information in quarterly model nowcasts simply by using Greenbook and mean SPF nowcasts to initialize model forecasts for future quarters.

The difference between using model versus expert nowcasts as initial conditions for model-based forecasts 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 beginning of the downturn than the model nowcasts. 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 forecasts. As a consequence, the model forecasts differ much less from each other than in the upper panel. The one-quarter-ahead model forecasts are more optimistic than the Greenbook. 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 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 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

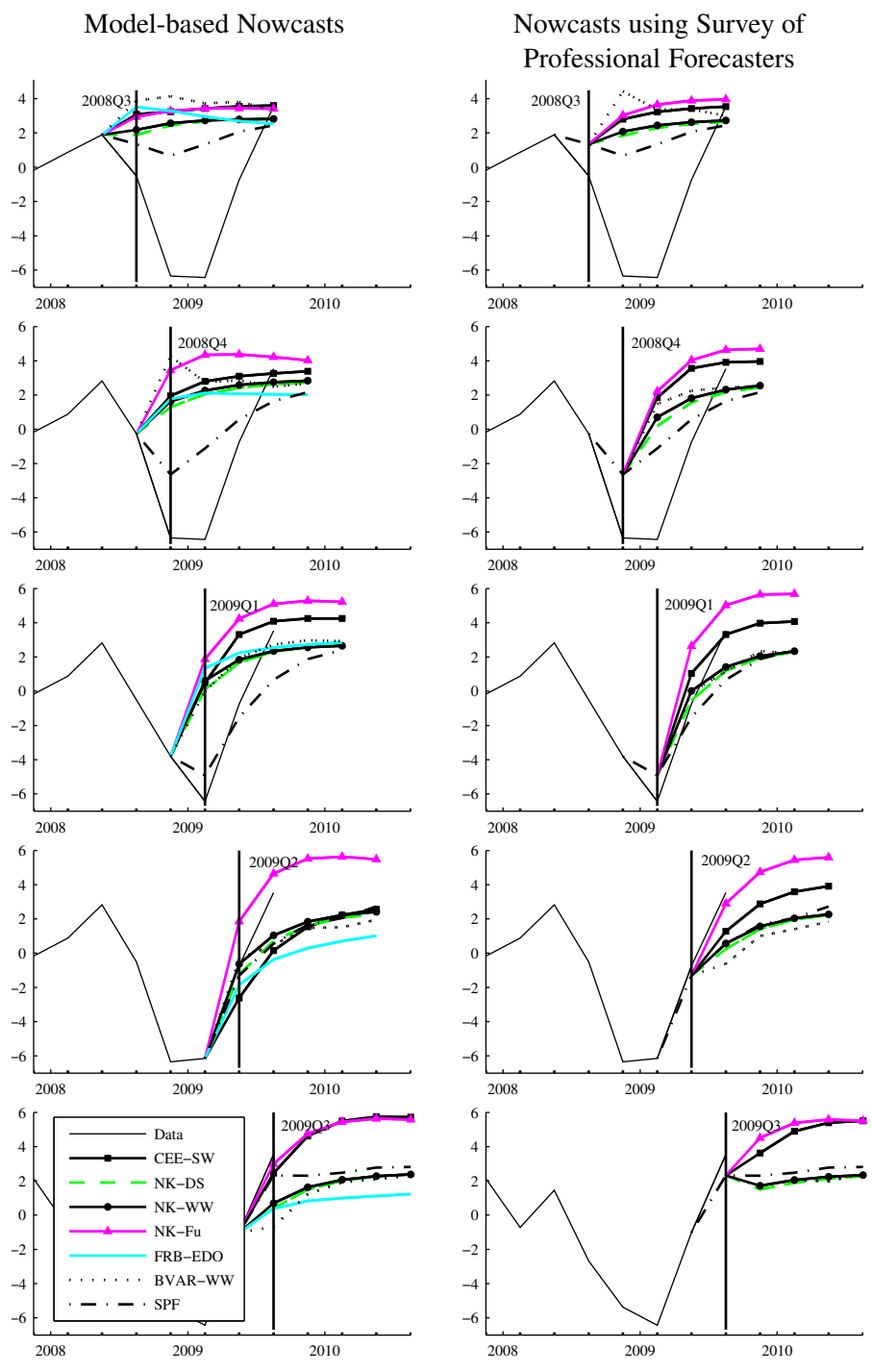


Figure 4: Real Output Growth forecast during the 2007-2009 recession (NBER defined peak: 2007Q4).

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.

with the mean SPF nowcasts further strengthens 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. Our typical recession sample covers the period from 4 quarters prior to the trough determined by the NBER Business Cycle Dating Committee to 4 quarters after the trough.¹¹ 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:

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

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 detailed comparison reveals some systematic differences. The CEE-SW model and the FRB/EDO model deliver fairly good forecasts in four out of five recessions. Several times, they yield the most accurate forecasts. In those cases where they are less precise than other models, the differences to the most accurate forecast are small. Both models have a rich economic structure and consider more observable data series than the other models. At the same time the parameterization is tight enough to yield accurate forecasts. The BVAR-WW model forecasts quite accurately in the

¹¹Exceptions are the 1980 and 2008/9 recessions. In the 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.

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 the 2001 recession, but more poorly in the other three recessions. Output growth in the 1990 and 2001 recession was less volatile. Perhaps, the lag structure of the Bayesian VAR is more appropriate during normal times and minor recessions. In more volatile times, sharp spikes in output fluctuations continue to feed through to forecasts for several quarters due to the lags included in the model. This results in less accurate forecasts.

The NK-DS and NK-WW models perform quite well during the most recent three recessions, but more poorly in the first two recessions. These models rely on three time series only. Persistence in output fluctuations arises primarily due to ad-hoc AR(1) shock processes. It is less pronounced than in the BVAR-WW model with four lags of endogenous variables. In these models a sharp spike in real GDP growth has a short but strong effect on the forecast. Finally, the NK-Fu model performs worse than the NK-DS and NK-WW models in most of the recessions. This model does not allow ad-hoc persistence via AR(1) shock processes. Shocks are assumed i.i.d. and output and inflation persistence can only arise from lags of output and inflation in the IS-curve and the overlapping wage structure. These dynamics may not be sufficient to yield precise output growth forecasts.

The mean model forecast shown in the seventh column of the table averages the six model forecasts. It performs very well. Most of the time it turns out to be fairly close to the best individual

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

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. As in the case of output growth it is difficult to draw general conclusions about how differences in models influence the forecasting results. The BVAR-WW yields very good forecasts for the three latest recessions, but performs worse for the two recessions in the 1980s. The reason might be that the BVAR-WW has a high a number of lags relative to the other models which may be more useful during less volatile times than during the 1980s disinflation. The CEE-SW model delivers one of the best inflation forecasts in several recessions and never one of the worst forecasts. In contrast to our findings for output growth, the FRB-EDO medium-scale model does not always

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

perform as well as CEE-SW in inflation forecasting. It delivers very good inflation forecasts in two of the five recessions, but is among the most inaccurate for the others. The NK-WW model performs better than the fairly similar NK-DS model, because the additional mark-up shock appear to better capture inflation dynamics. Finally, the NK-Fu model yields less satisfactory inflation forecasts. Perhaps, the overlapping wage contracts help the model capture the output-inflation tradeoff apparent in the 1980s recession but may induce more rigidity than required to match inflation dynamics in more recent recessions. 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 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

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

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 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 forecasting inflation is driven either by

modeling assumptions or information regarding the FOMC's objectives and future policies.

7 The Heterogeneity of Model-Based and Expert Forecasts

The model-based forecasts of output growth in the 2001 and 2008/09 recessions shown in Figures 1 to 4 indicate a substantial degree of heterogeneity that varies over time during these episodes. In this section, we 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:

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

where I_t^m 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 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 indicates the diversity of model forecasts while the solid line measures 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 models provides an indication of the extent of disagreement that may arise from different modeling assumptions, information sets and estimation methods. Since experts are faced with those same choices in developing their forecasting frameworks, the observed extent of heterogeneity in expert forecasts need not attributed to irrationality on behalf of individual forecasters.

We conduct some robustness checks to find out whether the heterogeneity measured by the standard deviation is strongly influenced by outliers. To this end, we compute the range between the 0.166 and 0.833 quantile for model-based and professional forecasts, that is 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 substantially over time. 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 several 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). Also, heterogeneity increases throughout the latter part of the 2008/09 recession for model as well as expert forecasts. Of course, 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. Such occasional spikes are not too surprising given that the SPF contains some extreme outliers. Rather, the co-movement visible in several episodes constitutes the more interesting finding, in our view.

Another aspect of heterogeneity concerns the range of accuracy of forecasts by individual fore-

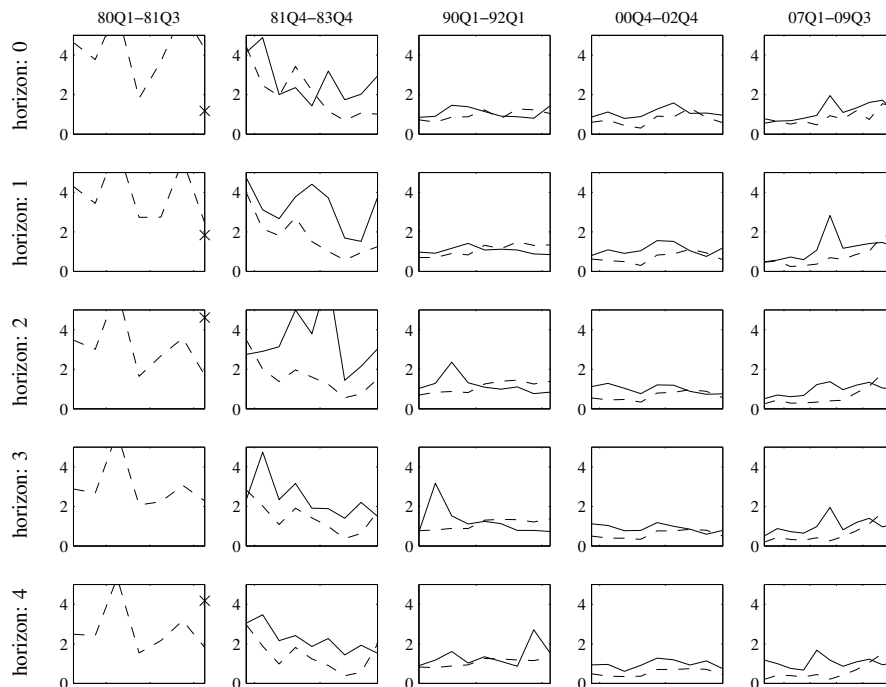


Figure 5: Standard Deviations of Output Growth Forecasts: Experts (solid) and Models (dashed)

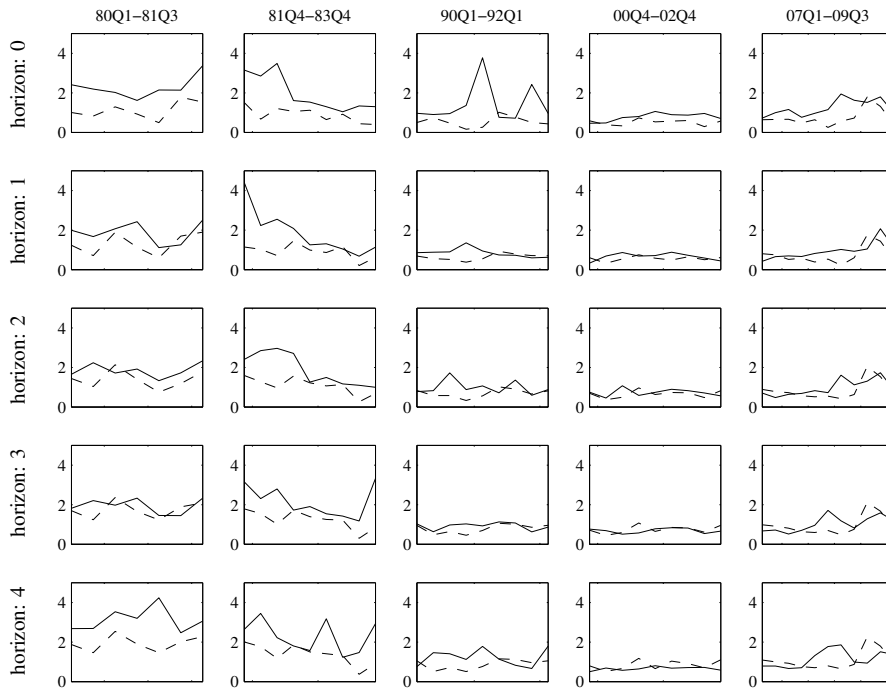


Figure 6: Standard Deviations of Inflation Forecasts: Survey (Black) and Models (Red)

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

Table 7 reports the worst, best and 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 lies 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 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 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

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

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

statistics. The models' minimal, maximal, and average RMSE decrease by a small amount.

8 Conclusions

In recent years, researchers such as Smets and Wouters (2004), Adolfson et al. (2005), Smets and Wouters (2007), Christoffel et al. (2008), Del Negro et al. (2007) and Wang (2009) have reported encouraging findings regarding the forecasting performance of state-of-the art structural models. By contrast, the failure of researchers and professional forecasters to predict the "Great Recession" of 2008 and 2009 has generated much public criticism regarding the state of economic forecasting and macroeconomic modeling. Against this background, our analysis of the forecasting performance of

models and experts during recessions provides several new insights.

The relative accuracy of model versus expert forecasts

First, we depart from the above-mentioned studies by using the real-time data vintages that were available in the past as the basis for evaluating forecasts of structural macroeconomic models. In doing so, we follow Faust and Wright (2009) who have shown that forecasts from non-structural models using ex-post revised data have uniformly smaller RMSEs than their counterparts estimated on real-time data. Thus, a comparison of structural model forecasts with historical expert forecasts has to be conducted on the basis of the real-time data vintages that could have been used by these experts at the time.¹²

Our focus on forecasting performance during recessions helps reveal that both, model and expert forecasts, tend to miss downturns. Interestingly, however, the model-based forecasts can do quite well during the recovery phase, sometimes even better than the Greenbook or mean-professional forecasts. Some model forecasts also predict the speed of recovery from the "Great Recession" surprisingly well. Model-based forecasts, in particular the mean model forecast¹³, compare quite well to the Greenbook and mean SPF forecasts, especially at a horizon of three to four quarters into the future. Overall, model-based forecasts still exhibit somewhat greater errors than expert forecasts, but this difference is surprisingly small considering that the models only take into account few economic variables and incorporate theoretical restrictions that are essential for evaluations of the impact of alternative policies but often considered a hindrance for effective forecasting.

Professional forecasters 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 their forecasts detect recessions a little earlier than model forecasts. However, model forecasts could be combined with such higher-frequency information (e.g. Giannone, Monti and Reichlin (2009)). To approximate the effect of efficient nowcasting we also conduct our comparisons between model-based and professional forecasts by starting from the professional nowcast. As a result, the gap between the two types of forecasts is further reduced.

Comparing model and expert forecast heterogeneity

We also quantify the extent of heterogeneity by means of the standard deviation across individual

¹²Faust and Wright (2009) find that the relative performance of non-structural models is less affected by using ex-post revised data. Whether this is also true for structural model still needs to be investigated.

¹³Our mean model forecast combines five structural models with a non-structural Bayesian VAR model. In light of the finding by Del Negro et al. (2007) that a 'hybrid' model which contains priors from a DSGE model and has otherwise a VAR structure performs better than either a structural DSGE model or a non-structural VAR this combination should be expected to improve forecast performance.

expert and model forecasts for a given forecasting horizon. The six model forecasts exhibit a broadly similar extent of forecast heterogeneity as the Survey of Professional Forecasters. 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. In some episodes the dynamics of forecast diversity derived from the two types of forecasts are quite similar.

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

How can the comparison of expert and model forecast heterogeneity be interpreted? Of course, some of the models considered were not available to professional forecasters during the earlier recession episodes. For example, state-of-the-art medium-scale DSGE models such as the CEE-SW and FRB-EDO models only became available in time for the recession of 2008/2009. Non-structural VAR models, however, have been used during all the episodes that we consider and the model of Fuhrer (1997) is representative of the New-Keynesian structural models that were already in use in the late 1980s and early 1990s. Furthermore, the reduced-form three-equation VAR implied by the linearized New-Keynesian models with microeconomic foundations (NK-DS and NK-WW) is not that different from the reduced-form VAR's implied by the earlier generation of New-Keynesian models. The microeconomic foundations simply imply additional cross-equation restrictions.

We interpret the comparison of the extent and dynamics of heterogeneity of model and expert forecasts as follows: while we can only speculate about the sources of disagreement among expert forecasters, the extent of disagreement among our six model forecasts can be traced to differences in modeling assumptions, different data coverage and different estimation methods. These three sources of disagreement are found to be sufficient to generate an extent of heterogeneity that is similar to the heterogeneity observed among expert forecasts. Furthermore, the recursive updating of model parameter estimates with incoming data induces dynamics in model forecast heterogeneity. In several episodes, expert forecast diversity even exhibits roughly similar variations. As a consequence of these findings, we would argue that it is not necessary to take recourse to irrational behavior or perverse incentives in order to explain the dynamics of expert forecast diversity.¹⁴ Rather, this diversity may largely be due to model uncertainty and belief updating in a world where the length of useful data

¹⁴Notwithstanding forecasters may face incentives to publish a forecast close to the consensus (Scharfstein and Stein, 1990; Lamont, 2002) or a very distinct forecast (Laster et al., 1999).

series is limited by structural breaks.¹⁵

On one side, our findings are encouraging in terms of the accuracy of forecasts derived from currently available structural macroeconomic models relative to expert forecasts from surveys. On the other side, our findings underscore the importance of research on models with heterogeneous expectations. Using models with homogeneous rational expectations for real-world forecasting, we estimate a significant range of forecast diversity that arises from different beliefs about appropriate modeling assumptions, estimation techniques and parameter estimates. This belief diversity itself may be a source of volatility. Of course, our models would attribute such volatility to shocks or other propagation mechanisms rather than endogenous heterogeneity in beliefs. Models with heterogeneous expectations provide an avenue for distinguishing this source of economic fluctuations from other candidate propagation mechanisms.

Clearly, this is an important area for research on macroeconomic modeling. One direction for progress is suggested by the theory of rational beliefs (see Kurz, 2009, for a detailed introduction into the theory of rational beliefs). Our set of models might be interpretable as beliefs in such a context. The theory of rational beliefs assumes people optimize given the limited knowledge they have and may make mistakes. They know that it is impossible to ever learn the true structural relationships and probability laws because structural breaks limit the length of useful data series. Diversity arises when market participants have different beliefs about the true data generating process and therefore estimate different models to forecast macroeconomic variables. Diverse beliefs are rational if they are consistent with the empirical distribution. The papers by Kurz and Motolese (2011), Guo et al. (2011) and Nielsen (2011) in this issue apply the theory of rational beliefs. Branch and McGough (2011), Branch and Evans (2011) and De Grauwe (2011) provide another avenue for studying heterogeneity of beliefs by modeling agents with cognitive limitations that generate boundedly rational forecasting rules. The latter two papers impose heterogeneous expectations directly into a New-Keynesian model. Instead of having rational expectations agents use small forecasting models. An interesting area for future research would be to estimate such models with heterogeneous expectations and compare the importance of belief diversity as a source of economic fluctuations relative to the propagation mechanisms considered by the homogeneous rational expectations models in our paper.

¹⁵Others have documented the strong time variation of disagreement among survey forecasts. For example, Mankiw et al. (2004) have investigated disagreement in inflation surveys. Engelberg et al. (2009) and Clements (2010) investigate the properties of SPF forecasts, the extent of heterogeneity and the cross-sectional histograms of survey forecasts. Similar in spirit to our analysis, Williams (2004) used multiple non-structural time series model to quantify the extent of inflation forecast heterogeneity due to model uncertainty. He concludes that model uncertainty provides an intuitively more appealing description of the observed diversity of inflation expectations than staggered information updating as suggested by Mankiw and Reis (2007).

References

- Adolfson, M., Andersson, M. K., Linde, J., Villani, M., and Vredin, A. (2005). Modern forecasting models in action: Improving macroeconomic analyses at central banks. Sveriges Riksbank Working Paper No. 190.
- An, S. and Schorfheide, F. (2007). Bayesian analysis of DSGE models. *Econometric Reviews*, 26(2-4):113–172.
- Bernanke, B. S. and Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3):525–546.
- Branch, W. A. and Evans, G. W. (2011). Monetary policy with heterogeneous expectations. *Economic Theory*, this volume.
- Branch, W. A. and McGough, B. (2011). Business cycle amplification with heterogeneous expectations. *Economic Theory*, this volume.
- Brock, W. and Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22:1235–1274.
- Bryant, R., Currie, D., Frenkel, J., Masson, P., and Portes, R., editors (1989). *Macroeconomic Policies in an Interdependent World*. Washington, D.C.: The Brookings Institution.
- Bryant, R., Henderson, D. W., Holtham, G., Hooper, P., and Symansky, S. A., editors (1988). *Empirical Macroeconomics for Interdependent Economies*. Washington, D.C.: The Brookings Institution.
- Bryant, R., Hooper, P., and Mann, C., editors (1993). *Evaluating Policy Regimes: New Research in Empirical Macroeconomics*. Washington, D.C.: The Brookings Institution.
- Capistran, C. and Timmermann, A. (2009). Disagreement and biases in inflation expectations. *Journal of Money, Credit, and Banking*, 41:365–396.
- Chiarella, C., Dieci, R., and He, X.-Z. (2007). Heterogeneous expectations and speculative behavior in a dynamic multi-asset framework. *Journal of Economic Behavior and Organization*, 62:408–427.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy*, 113(1):1–45.
- Christoffel, K., Coenen, G., and 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.

- Clements, M. P. (2010). Explanations of the inconsistencies in survey respondents' forecasts. *European Economic Review*, 54(4):536–549.
- De Grauwe, P. (2011). Animal spirits and monetary policy. *Economic Theory*, this volume.
- Del Negro, M. and Schorfheide, F. (2004). Priors from general equilibrium models for VARs. *International Economic Review*, 45(2):643–673.
- Del Negro, M., Schorfheide, F., Smets, F., and Wouters, R. (2007). On the fit of New Keynesian models. *Journal of Business and Economic Statistics*, 25(2):123143.
- Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3:1 – 100.
- Edge, R. M., Kiley, M. T., and Laforge, 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., and Laforge, 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., and Laforge, 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*, forthcoming.
- Engelberg, J., Manski, C. F., and Williams, J. (2009). Assessing the temporal variation of macroeconomic forecasts by a panel of changing composition. Northwestern University Working Paper.
- Faust, J. and 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.
- Francis, N. and Ramey, V. A. (1995). Measures of per capita hours and their implications for the technology-hours debate. NBER Working Paper 11694.
- 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. and Moore, G. (1995a). Inflation persistence. *The Quarterly Journal of Economics*, 110(1):127–159.
- Fuhrer, J. C. and 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., and Reichlin, L. (2009). Incorporating conjunctural analysis in structural models. In Wieland, V., editor, *The science and practice of monetary policy today*, pages 41–57. Springer Science.
- Giordani, P. and Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, 47:1037–1059.
- Goodfriend, M. and King, R. G. (1997). The New Neoclassical Synthesis and the role of monetary policy. In Bernanke, B. S. and Rotemberg, J. J., editors, *National Bureau of Economic Research Macroeconomics Annual 1997*. MIT Press, Cambridge, MA.
- Guo, W. C., Wang, F. Y., and Wu, H. M. (2011). Financial leverage and market volatility with diverse beliefs. *Economic Theory*, this volume.
- Hamilton, J. D. (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.
- Klein, L., editor (1991). *Comparative Performance of U.S. Econometric Models*. Oxford, Eng.: Oxford University Press.
- 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. (1997a). Endogenous economic fluctuations and rational beliefs: A general perspective. In Kurz, M., editor, *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*. Springer Series in Economic Theory, No. 6, Springer Verlag.
- Kurz, M., editor (1997b). *Endogenous Economic Fluctuations: Studies in the Theory of Rational Beliefs*. Springer Series in Economic Theory, No. 6, Springer Verlag.
- Kurz, M. (2009). Rational diverse beliefs and market volatility. In Hens, T. and Schenk-Hoppe, K., editors, *Handbook of financial markets: dynamics and evolution*. North Holland.
- Kurz, M., Jin, H., and Motolese, M. (2003). *Knowledge, Information and Expectations in Modern Macroeconomics: Essays In Honor of Edmund S. Phelps*, chapter 10: Endogenous Fluctuations and the Role of Monetary Policy, pages 188–227. Princeton University Press: Princeton, N.J.
- Kurz, M., Jin, H., and Motolese, M. (2005). The role of expectations in economic fluctuations and

- the efficacy of monetary policy. *Journal of Economic Dynamics & Control*, 29:20172065.
- Kurz, M. and Motolesse, M. (2011). Diverse beliefs and time variability of risk premia. *Economic Theory*, this volume.
- Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior and Organization*, 48(3):265–280.
- Laster, D., Bennett, P., and Geoum, I. S. (1999). Rational bias in macroeconomic forecasts. *Quarterly Journal of Economics*, 114(1):293–318.
- Mankiw, N. G. and Reis, R. (2007). Sticky information in general equilibrium. *Journal of the European Economic Association*, 5 (2-3):603–613.
- Mankiw, N. G., Reis, R., and Wolfers, J. (2004). Disagreement about inflation expectations. In Gertler, M. and Rogoff, K., editors, *NBER Macroeconomics Annual 2003*, volume 18, pages 209–248. Cambridge, Mass.: MIT Press.
- Mishkin, F. (2004). Discussion of 'Disagreement about inflation expectations' by N. Gregory Mankiw, Ricardo Reis, and Justin Wolfers. In Gertler, M. and Rogoff, K., editors, *NBER Macroeconomics Annual 2003*, volume 18, pages 257–268. Cambridge, Mass.: MIT Press.
- Nielsen, C. K. (2011). Price stabilizing, pareto improving policies. *Economic Theory*, this volume.
- Romer, C. D. and Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 90:429–457.
- Rotemberg, J. J. and 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.
- Scharfstein, D. S. and Stein, J. C. (1990). Herd behavior and investment. *American Economic Review*, 80(3):465–479.
- 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. and 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. and Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *The American Economic Review*, 97(3):586–606.
- Taylor, J. B. (1979). Estimation and control of a macroeconomic model with rational expectations.

Econometrica, 47 (5):1267–1286.

- Taylor, J. B. (1993). *Macroeconomic Policy in a World Economy*. W.W. Norton, New York.
- Taylor, J. B. and Wieland, V. (2009). Surprising comparative properties of monetary models: Results from a new data base. NBER Working Paper 14849.
- Walsh, C. E. (2003). *Monetary theory and policy*. The MIT Press, Cambridge.
- 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., editor (2009). *The Science and Practice of Monetary Policy Today*. Springer Science.
- Wieland, V., Cwik, T., Müller, G. J., Schmidt, S., and Wolters, M. (2009). A new comparative approach to macroeconomic modeling and policy analysis. Manuscript, Center for Financial Studies, Frankfurt.
- Williams, J. C. (2004). Discussion of 'Disagreement about inflation expectations' by N. Gregory Mankiw, Ricardo Reis, and Justin Wolfers. In Gertler, M. and Rogoff, K., editors, *NBER Macroeconomics Annual 2003*, volume 18, pages 257–268. Cambridge, Mass.: MIT Press.
- Woodford, M. (2003). *Interest and prices: Foundations of a theory of monetary policy*. Princeton University Press, Princeton.

Appendix A1: The Macroeconomic Models Used to Compute Forecasts

This appendix provides a description of the six macroeconomic models that are used in this paper to generate forecasts. In the case of the NK-Fu, NK-DS, CEE-SW and FRB/EDO models our notation follows exactly the notation in the model authors' original articles.

BVAR-WW Model: Non-structural VAR models have been available to forecasters for decades and are still being used by practitioners today. Such a 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. We estimate such a VAR on output growth, inflation and the federal funds rate using Bayesian methods. Each of the variables is regressed on a constant, four lagged values of the variable itself and four lagged values of the other two variables. It is well known that unrestricted VARs are heavily overparameterized. To improve forecast performance it is important to shrink the parameter space in some manner. We follow Doan et al. (1984) and use the so-called Minnesota prior to avoid over-parameterization. This prior implies shrinking the parameters towards zero by assuming that the price level, real output and the interest rate follow independent random walks. All parameters are assumed to be normally distributed with mean zero. The variance around these zero priors decreases with lag-length. The rationale for this assumption is that short lags contain more information about the dependent variables than long lags.

NK-Fu Model: The model of Fuhrer (1997) is a good example of the New-Keynesian models that were developed in the 1980s and early 1990s.¹⁶ While academics still focused primarily on developing the microeconomic foundations of real business cycle theory, these models became quite popular among central bank researchers and practitioners. They took into account adaptive and forward-looking behavior of market participants, real effects of monetary policy and output and inflation persistence. The model of Fuhrer (1997) exhibits a high degree of inertia with respect to aggregate demand which is determined by the following IS-curve:

$$\tilde{y}_t = a_0 + a_1\tilde{y}_{t-1} + a_2\tilde{y}_{t-2} + a_\rho\rho_{t-1} + \varepsilon_{y,t}, \quad (13)$$

\tilde{y}_t denotes the output gap, which is computed as the deviation from the log-linear trend. ρ_t denotes the long-term real interest rate and $\varepsilon_{y,t}$ a demand shock. The long-term real interest rate is determined by an intertemporal arbitrage condition that equalizes the expected holding-period yields on government

¹⁶For other examples see the model comparison projects of Bryant et al. (1988), Bryant et al. (1989), Klein (1991), and Bryant et al. (1993).

bonds and real long-term bonds:

$$\rho_t - D[E_t(\rho_{t+1}) - \rho_t] = f_t - E_t(\pi_{t+1}). \quad (14)$$

f_t denotes the federal funds rate, π_t the quarterly inflation rate and D is a constant approximation for Macaulay's duration that is set equal to 10 years.

The short-run aggregate supply nexus between output and inflation is importantly influenced by overlapping wage contracts. Fuhrer assumes that wage contracts that remain in effect for one to four quarters are negotiated relative to the real wage implied by those set in the recent past and those that are expected to be negotiated in the near future (see Fuhrer and Moore, 1995a,b). v_t denotes an index of wage contracts that are currently in effect:

$$v_t = \sum_{i=0}^3 \omega_i (x_{t-i} - p_{t-i}), \quad (15)$$

where x_t denotes the log wage contract negotiated in period t and p_t the log price level. The weights ω_i are the proportions of the outstanding contracts and sum to one. The weights decrease for contracts negotiated in earlier periods. The current nominal wage contract is determined such that the current real wage contract equals the average real contract wage index expected to prevail over the life of the contract. Additionally, it adjusted for expected excess demand conditions as reflected in current and expected future output gaps:

$$x_t - p_t = \sum_{i=0}^3 \omega_i (v_{t+i} + \gamma \tilde{y}_{t+i}) + \varepsilon_{p,t}. \quad (16)$$

$\varepsilon_{p,t}$ is a cost-push shock. The aggregate log wage index is a weighted average of the log of wage contracts. The aggregate price level is a constant mark-up (normalized to zero) over the aggregate wage rate. Inflation dynamics depend on current, past and expected future demand. The model is quite successful in matching the strong inflation persistence observed in U.S. data. Inflation is given by an average of changes in the log nominal wage contracts:

$$\pi_t = \sum_{i=0}^3 \omega_i (x_{t-i} - x_{t-i-1}). \quad (17)$$

The model is closed with a monetary policy reaction function. The Fed is assumed to set the federal funds rate with respect to a constant equilibrium value, the lagged funds rate, inflation, lagged inflation, the output gap and the change in the output gap. Deviations from the reaction function are interpreted as monetary policy shocks:

$$f_t = \alpha_0 + \alpha_{f1} f_{t-1} + \alpha_{\pi 0} \pi_t + \alpha_{\pi 1} \pi_{t-1} + \alpha_{\Delta y} (\tilde{y}_t - \tilde{y}_{t-1}) + \alpha_y \tilde{y}_t + \varepsilon_{f,t}. \quad (18)$$

Contrary to the other structural models considered in this paper, Fuhrer allows for the possibility of contemporaneously correlated structural shocks. The variance-covariance matrix is estimated together with the parameters of the model.

NK-DS Model: The model by Del Negro and Schorfheide (2004) is an example of small-scale New-Keynesian models with microeconomic foundations in the vein of Rotemberg and Woodford (1997) and Goodfriend and King (1997). A representative household derives utility from consumption relative to a habit stock that depends on the level of technology. Hours worked reduce the household's utility and real money balances increase it. The utility function is additively separable. Utility is maximized over an infinite lifetime subject to the household's budget constraint. The household earns income from different sources: wage income from supplying perfectly elastic labor services to firms, interest rate payments from bond holdings and profits from the firms. It pays lump-sum taxes. Utility maximization implies an Euler equation. Linearizing this equation and imposing market clearing (output equals consumption and government spending) yields the New-Keynesian forward-looking IS-equation:

$$x_t = E_t x_{t+1} - \tau^{-1}(R_t - E_t \pi_{t+1}) + (1 - \rho_g)g_t + \rho_z \tau^{-1} z_t, \quad (19)$$

x_t denotes output, π_t inflation and R_t the federal funds rate. τ is the risk aversion parameter of the household. All variables are defined in percentage deviations from steady state. g_t and z_t are government spending and technology shock processes. Both shocks follow AR(1) processes (not shown) with parameters ρ_g and ρ_z . The government consumes a fraction of output which fluctuates exogenously according to the shock process: ξ_t denotes the fraction of output consumed by the government and the shock is defined as $g_t = 1/(1 - \xi_t)$. The government issues bonds that can be bought by households and it collects lump-sum taxes to finance its expenditures.

The production sector consists of a continuum of monopolistically competitive firms that are owned by the households. They face demand curves that can be derived from a Dixit-Stiglitz final good aggregator. Nominal rigidities are modelled via quadratic price adjustment costs. Firms pay these costs in form of an output loss when they desire to set a price in deviation from the level implied by steady-state inflation. The production function is linear in labor. Labor is hired from the households. Total factor productivity follows a unit root process. Thus, it induces a stochastic trend into the model. As a result, output fluctuates around the steady-state growth rate. Firms maximize the present value of expected profits over an infinite horizon. The optimality condition implies that prices are set as a fixed mark-up over marginal cost. Linearizing this first order condition leads to the

following New-Keynesian forward-looking Phillips curve:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa(x_t - g_t), \quad (20)$$

where β is the household's discount factor and κ is a function of the price adjustment cost parameter and the elasticity of demand. Inflation is a function of marginal cost which can be substituted with the output gap. The model is closed with a monetary policy rule. The rule assumes that the central bank sets the current interest rate as a function of current inflation, the output gap, and the previous interest rate choice:

$$R_t = \rho_R R_{t-1} + (1 - \rho_R)(\psi_1 \pi_t + \psi_2 x_t) + \varepsilon_{R,t}. \quad (21)$$

The monetary policy shock, $\varepsilon_{R,t}$, is assumed iid-normally distributed. ρ_R indicates the degree of interest rate smoothing and ψ_1 and ψ_2 capture the policy response to inflation and output gaps. The IS equation and the policy rule together represent the aggregate demand side, while the Phillips curve captures fluctuations in aggregate supply.

NK-WW model: The NK-WW model generalizes the NK-DS model in terms of the economic shocks considered. To allow for richer output and inflation dynamics we add serially correlated preference and mark-up shock processes χ_t and Φ_t . Both shocks follow AR(1) processes with parameters ρ_χ and ρ_Φ . The preference shock enters the consumption term in the utility function and appears in the New-Keynesian IS-equation:

$$x_t = E_t x_{t+1} - \tau^{-1}(R_t - E_t \pi_{t+1}) + (1 - \rho_g)g_t + \rho_z \tau^{-1} z_t + \tau^{-1}(1 - \rho_\chi)\chi_t, \quad (22)$$

Both shocks enter the New-Keynesian Phillips curve. The mark-up shock has a direct effect on inflation. The preference shock influences marginal costs and thereby also inflation determination:

$$\pi_t = \beta E_t \pi_{t+1} + \kappa [x_t - g_t + \tau^{-1}(\Phi_t - \chi_t)]. \quad (23)$$

The monetary policy rule is the same as in the NK-DS model.

CEE-SW Model: Building on the above-mentioned micro-founded New-Keynesian model Christiano, Eichenbaum and Evans (2005) developed the first medium-scale New-Keynesian DSGE (dynamic stochastic general equilibrium) model that can fit a significant number of important empirical regularities of the U.S. economy (NBER working paper 2001). Smets and Wouters (2003, 2007) extended this model and estimated it with Bayesian methods. The CEE-SW model contains a large number of frictions and structural shocks. Physical capital is included in the production function

and capital formation is endogenous. 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. Utility is nonseparable in consumption and leisure. There exist fixed costs in production and the Dixit-Stiglitz aggregator is replaced with the aggregator by Kimball (1995) which implies a non-constant elasticity of demand. The model contains seven structural shocks and is fit to seven time series. Among the shocks are, total factor productivity, risk premium, investment-specific technology, wage mark-up, price mark-up, government spending and monetary policy shocks. All shock processes are serially correlated. In the following we describe each of the linearized equations of the model following the notation in Smets and Wouters (2007).

The resource constraint is given by:

$$y_t = c_y c_t + i_y i_t + z_y z_t + \varepsilon_t^g, \quad (24)$$

where output y_t is the sum of consumption, c_t , and investment, i_t , weighted with their steady state ratios to output (c_y and i_y), the capital-utilization cost which depends on the capital utilization rate, z_t , and an exogenous government spending shock ε_t^g . ε_t^g follows an AR(1) process and is also affected by the technology shock. z_y equals $R_*^k k_y$, where k_y is the ratio of capital to output in steady state and R_*^k is the rental rate of capital in steady state. Combining the households' first order conditions for consumption and bond holdings yields the consumption Euler equation

$$c_t = c_1 c_{t-1} + (1 - c_1) E_t(c_{t+1}) + c_2 (l_t - E_t(l_{t+1})) - c_3 (r_t - E_t(\pi_{t+1})) + \varepsilon_t^b. \quad (25)$$

The parameters are $c_1 = (\lambda/\gamma)/(1 + \lambda/\gamma)$, $c_2 = [(\sigma_c - 1)(W_*^h L_*/C_*)]/[(\sigma_c(1 + \lambda/\gamma)]$ and $c_3 = (1 - \lambda/\gamma)/[(1 + \lambda/\gamma)\sigma_c]$. λ governs the degree of habit formation, γ is the labor augmented steady growth rate, σ_c the inverse of the intertemporal elasticity of substitution and parameters with a * subscript denote steady state values. ε_t^b denotes an AR(1) shock process on the premium over the central bank controlled interest rate. Consumption is a weighted average of past and expected consumption due to habit formation. The consumption Euler equation depends on hours worked, l_t , because of the nonseparability of utility. When consumption and hours are complements ($\sigma_c > 1$), consumption increases with current hours and decreases with expected hours next period. The real interest rate and the shock term affect aggregate demand by inducing intertemporal substitution in consumption.

The investment Euler equation is given by

$$i_t = i_1 i_{t-1} + (1 - i_1) E_t(i_{t+1}) + i_2 q_t + \varepsilon_t^i, \quad (26)$$

where $i_1 = 1/(1 + \beta\gamma^{1-\sigma_c})$ and $i_2 = [1/(1 + \beta\gamma^{1-\sigma_c})\gamma^2\phi]$. β denotes the discount factor, ϕ the elasticity of the capital adjustment cost function, q_t Tobin's Q and ε_t^i an investment specific technology

shock that follows an AR(1) process. Current investment is a weighted average of past and expected future investment due to the existence of capital adjustment costs. It is positively related to the real value of the existing capital stock. This dependence decreases with the elasticity of the capital adjustment cost function. The arbitrage equation for the real value of the capital stock is:

$$q_t = q_1 E_t(q_{t+1}) + (1 - q_1) E_t(r_{t+1}^k) - (r_t - E_t(\pi_{t+1}) + \varepsilon_t^b), \quad (27)$$

where $q_1 = \beta \gamma^{-\sigma_c} (1 - \delta)$. r_t^k denotes the real rental rate of capital and δ the depreciation rate of capital. The real value of the existing capital stock is a positive function of its expected value next period and the rental rate on capital and a negative function of the real interest rate and the external finance premium.

The production process is assumed to be determined by a Cobb-Douglas production function with fixed costs:

$$y_t = \phi_p (\alpha k_t^s + (1 - \alpha) l_t + \varepsilon_t^a). \quad (28)$$

k_t^s denotes effective capital (physical capital adjusted for the capital utilization rate), ε_t^a a neutral productivity shock that follows an AR(1) process and ϕ_p is one plus the share of fixed costs in production. Output is produced using capital and labour and is boosted by technology shocks. Capital used in production depends on the capital utilization rate and the physical capital stock of the previous period as new capital becomes effective with a lag of one quarter:

$$k_t^s = k_{t-1} + z_t. \quad (29)$$

Household income from renting capital services to firms depends on r_t^k and changing capital utilization is costly so that the capital utilization rate depends positively on the rental rate of capital:

$$z_t = (1 - \psi) / \psi r_t^k, \quad (30)$$

where $\psi \in [0, 1]$ is a positive function of the elasticity of the capital utilization adjustment cost function. The law of motion for physical capital is given by:

$$k_t = k_1 k_{t-1} + (1 - k_1) i_t + k_2 \varepsilon_t^i, \quad (31)$$

where $k_1 = (1 - \delta) / \gamma$ and $k_2 = (1 - (1 - \delta) / \gamma) (1 + \beta \gamma^{1 - \sigma_c}) \gamma^2 \phi$. The price mark-up μ_t^p equals the difference between the marginal product of labor and the real wage w_t :

$$\mu_t^p = \alpha (k_t^s - l_t) + \varepsilon_t^a - w_t. \quad (32)$$

Monopolistic competition, Calvo-style price contracts, and indexation of prices that are not free to be chosen optimally combine to yield the following Phillips curve:

$$\pi_t = \pi_1 \pi_{t-1} + \pi_2 \pi_{t+1} - \pi_3 \mu_t^p + \varepsilon_t^p, \quad (33)$$

with $\pi_1 = \iota_p / (1 + \beta \gamma^{1-\sigma_c} \iota_p)$, $\pi_2 = \beta \gamma^{1-\sigma_c} / (1 + \beta \gamma^{1-\sigma_c} \iota_p)$, and $\pi_3 = 1 / (1 + \beta \gamma^{1-\sigma_c} \iota_p) [(1 - \beta \gamma^{1-\sigma_c} \xi_p)(1 - \xi_p) / \xi_p ((\phi_p - 1)\varepsilon_p + 1)]$. This Phillips curve contains not only a forward-looking but also a backward-looking inflation term because of price indexation. Firms that cannot adjust prices optimally either index their price to the lagged inflation rate or to the steady-state inflation rate. Note, this indexation assumption ensures also that the long-run Phillips curve is vertical. ξ_p denotes the Calvo parameter, ι_p governs the degree of backward indexation, ε_p determines the curvature of the Kimball (1995) aggregator. The Kimball aggregator complementarity effects enhance the price rigidity resulting from Calvo-style contracts. The mark-up shock ε_t^p follows an ARMA(1,1) process. A monopolistic labor market yields the condition that the wage mark-up μ_t^w equals the real wage minus the marginal rate of substitution mrs_t :

$$\mu_t^w = w_t - mrs_t = w_t - (\sigma_l l_t + \frac{1}{1 - \lambda/\gamma} (c_t - \lambda/\gamma c_{t-1})), \quad (34)$$

with σ_l being the Frisch elasticity of labor supply. The wage Phillips-Curve is given by:

$$w_t = w_1 w_{t-1} + (1 - w_1)(E_t(w_{t+1}) + E_t(\pi_{t+1})) - w_2 \pi_t + w_3 \pi_{t-1} - w_4 \mu_t^w + \varepsilon_t^w, \quad (35)$$

where $w_1 = 1 / (1 + \beta \gamma^{1-\sigma_c})$, $w_2 = (1 + \beta \gamma^{1-\sigma_c} \iota_w) / ((1 + \beta \gamma^{1-\sigma_c}))$, $w_3 = \iota_w / (1 + \beta \gamma^{1-\sigma_c})$, and $w_4 = 1 / (1 + \beta \gamma^{1-\sigma_c}) [(1 - \beta \gamma^{1-\sigma_c} \xi_w)(1 - \xi_w) / (\xi_w ((\phi_w - 1)\varepsilon_w + 1))]$. The parameter definition is analogous to the price Phillips curve.

Setting $\xi_p = 0$, $\xi_w = 0$, $\varepsilon_t^p = 0$ and $\varepsilon_t^w = 0$ one obtains the efficient flexible price and flexible wage allocation. The output gap x_t is defined as the log difference between output and flexible price output just like in the small-scale New-Keynesian models above.

The monetary policy rule reacts to inflation, the output gap and the change in the output gap and incorporates partial adjustment:

$$r_t = \rho r_{t-1} + (1 - \rho)(r_\pi \pi_t + r_x x_t) + r_{\Delta x_t} (x_t - x_{t-1}) + \varepsilon_t^r. \quad (36)$$

ε_t^r is a monetary policy shock that follows an AR(1) process.

FRB-EDO Model: The model by Edge et al. (2008) is a more disaggregated DSGE model that was developed at the Board of Governors of the Federal Reserve System. 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 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. 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 to fit 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. We estimate a variant of the FRB-EDO model that is built as close to the documentation of (Edge et al., 2007) as possible. While the aggregate dynamics implied by our version of the model do not exactly match the figures in the authors' documentation, they come reasonably close to that.

In the following we describe the main equations of the model. There are two types of intermediate-good producing firms that differ with respect to the rate of technological progress in their production technology. Production depends on technology, utilized non-residential capital and labor. Non-residential capital is rented from capital owners and labor is hired from households. The first sector is called the business and institutions sector and most of its output is used for consumption. The sector is therefore denoted by *cbi*. The technology of the second sector grows at a faster rate. This sector is called the business sector and the produced goods are used for capital accumulation. It is therefore denoted by *kb*.

The intermediate-goods producing firms' cost minimization problems with respect to labor and utilized non-residential capital lead to the following optimal factor input conditions:

$$L_t^s = (1 - \alpha) \tilde{X}_t^s \frac{\tilde{MC}_t^s}{\tilde{W}_t^s}, \quad \text{for } s = cbi, kb \quad (37)$$

$$\frac{\tilde{K}_t^{u,nr,s}}{\Gamma_t^{x,kb}} = \alpha \tilde{X}_t^s \frac{\tilde{MC}_t^s}{\tilde{R}_t^{nr,s}}, \quad \text{for } s = cbi, kb \quad (38)$$

L_t^s is the labor input, \tilde{X}_t^s are the produced goods, \tilde{MC}_t^s are marginal costs, \tilde{W}_t^s is the nominal wage rate, $\tilde{K}_t^{u,nr,s}$ is the amount of utilized non-residential capital, $\Gamma_t^{x,kb}$ is the growth rate of output in the *kb* sector, $\tilde{R}_t^{nr,s}$ is the aggregate nominal rental rate on non-residential capital and α denotes the capital share in the production function. A tilde on a variable denotes stationarized variables.

The stationarized production function is given by:

$$\tilde{X}_t^s = (L_t^s)^{(1-\alpha)} \left(\frac{\tilde{K}_t^{u,nr,s}}{\Gamma_t^{x,kb}} \right)^\alpha, \quad \text{for } s = cbi, kb \quad (39)$$

The intermediate-goods firms face monopolistic competition. Thus, they are able to set prices that maximize the present value of profits in the infinite future. When maximizing profits the firms have to take into account the demand for their goods. This demand function is derived from perfectly competitive final good firms that use a Dixit-Stiglitz aggregation function. Furthermore, price adjustment is constrained by a quadratic adjustment cost function. Adjustment costs are paid in the form of an

output loss when the price adjustment exceeds an average of the steady state inflation rate and last period's inflation rate. The Phillips curve is given by:

$$\begin{aligned} \Theta_t^{x,s} \widetilde{MC}_t^s \widetilde{X}_t^s &= (\Theta_t^{x,s} - 1) \widetilde{P}_t^s \widetilde{X}_t^s \\ &+ 100\chi^p (\Pi_t^{p,s} - \eta^p \Pi_{t-1}^{p,s} - (1 - \eta^p) \Pi_*^{p,s}) \Pi_t^{p,s} \widetilde{X}_t^s \widetilde{P}_t^s \\ &- \beta E_t \left(\frac{\widetilde{\Lambda}_{t+1}^{cnn}}{\widetilde{\Lambda}_t^{cnn}} 100\chi^p (\Pi_{t+1}^{p,s} - \eta^p \Pi_t^{p,s} - (1 - \eta^p) \Pi_*^{p,s}) \Pi_{t+1}^{p,s} \widetilde{P}_{t+1}^s \widetilde{X}_{t+1}^s \right), \end{aligned} \quad (40)$$

where $s = cbi, kb$. $\Theta_t^{x,s}$ is the stochastic elasticity of substitution between differentiated intermediate goods and governs shocks to the price mark-up over marginal cost. $\Pi_t^{p,s}$ is the inflation rate and $\Pi_*^{p,s}$ is the steady state inflation rate. \widetilde{P}_t^s is the price level relative to the *cbi* sector (\widetilde{P}_t^{cbi} is equal to 1). $\widetilde{\Lambda}_t^{cnn}$ denotes the marginal utility of the consumption good. The parameter χ^p reflects the size of adjustment costs in re-setting prices. η^p determines the relative importance of lagged inflation and steady state inflation in the adjustment cost function and β is the household's discount factor.

There are three different types of capital owners who invest in goods, transform these into the three different capital stocks and rent them to households and firms. Goods from the fast growing sector (kb) are transformed into non-residential capital or consumer-durable capital. Goods from the slow growing sector (cbi) are transformed into residential capital stock or directly used for household consumption. Capital evolution depends on a quadratic investment adjustment cost that is paid via a capital loss if current investment differs from investment in the previous period adjusted by the growth rate of the respective sector production. In addition there are stochastic capital efficiency shocks. The first-order condition of the non-residential capital owners with respect to the capital stock is given by:

$$\widetilde{Q}_t^{nr} = \beta E_t \left\{ \frac{\widetilde{\Lambda}_{t+1}^{cnn}}{\widetilde{\Lambda}_t^{cnn}} \frac{1}{\Gamma_{t+1}^{x,kb}} \left(\widetilde{R}_{t+1}^{nr} + (1 - \delta^{nr}) \widetilde{Q}_{t+1}^{nr} \right) \right\}, \quad (41)$$

where \widetilde{Q}_t^{nr} is the price of installed non-residential capital, \widetilde{R}_t^{nr} is the nominal rental rate on non-residential capital and δ^{nr} is the depreciation rate. The first order condition with respect to investment in non-residential capital is given by:

$$\begin{aligned} \widetilde{P}_t^{kb} &= \widetilde{Q}_t^{nr} \left[A_t^{nr} - 100\chi^{nr} \left(\frac{\widetilde{E}_t^{nr} - \widetilde{E}_{t-1}^{nr}}{\widetilde{K}_t^{nr}} \Gamma_t^{x,kb} \right) \right] \\ &+ \beta E_t \left\{ \frac{\widetilde{\Lambda}_{t+1}^{cnn}}{\widetilde{\Lambda}_t^{cnn}} \widetilde{Q}_{t+1}^{nr} 100\chi^{nr} \left(\frac{\widetilde{E}_{t+1}^{nr} - \widetilde{E}_t^{nr}}{\widetilde{K}_{t+1}^{nr}} \Gamma_{t+1}^{x,kb} \right) \right\}. \end{aligned} \quad (42)$$

A_t^{nr} is a capital efficiency shock, χ^{nr} is an investment adjustment cost parameter, \widetilde{E}_t^{nr} denotes expenditure on goods used for non-residential investment and \widetilde{K}_t^{nr} is the non-residential capital stock. Other conditions that include the capital accumulation equation and the market clearing condition for

non-residential capital used in the production process in both sectors are given by:

$$\tilde{R}_t^{nr,s} = \frac{\tilde{R}_t^{nr}}{U_t^s}, \quad \text{for } = cbi, kb \quad (43)$$

$$U_t^s = \left(\frac{1}{\kappa} \frac{\tilde{R}_t^{nr,s}}{\tilde{P}_t^{kb}} \right)^{\frac{1}{\psi}}, \quad \text{for } = cbi, kb \quad (44)$$

$$\tilde{K}_{t+1}^{nr} = (1 - \delta^{nr}) \frac{\tilde{K}_t^{nr}}{\Gamma_t^{x,kb}} + A_t^{nr} \tilde{E}_t^{nr} - \frac{100\chi^{nr}}{2} \left(\frac{\tilde{E}_t^{nr} - \tilde{E}_{t-1}^{nr}}{\tilde{K}_t^{nr}} \Gamma_t^{x,kb} \right)^2 \frac{\tilde{K}_t^{nr}}{\Gamma_t^{x,kb}} \quad (45)$$

$$\tilde{K}_t^{nr} = \tilde{K}_t^{nr,cbi} + \tilde{K}_t^{nr,kb}. \quad (46)$$

U_t^s is the capital utilization rate, κ is a scaling parameter for the cost of changing the capacity utilization rate and ψ is the elasticity of the capacity utilization cost. $\tilde{R}_t^{nr,cbi}$ and $\tilde{R}_t^{nr,kb}$ denote the nominal rental rate on non-residential capital used in the cbi and kb sector denoted by $\tilde{K}_t^{nr,cbi}$ and $\tilde{K}_t^{nr,kb}$, respectively.

The first order conditions for the consumer durable capital owners and residential capital owners are similar. As these types of capital are not used in the production process, there are only three first order conditions for each capital owner. The only difference between the two types of capital is that the consumer durable capital good is produced in the fast growing (kb) sector and the residential capital good is produced in the slow growing (cbi) sector:

$$\tilde{Q}_t^{cd} = \beta E_t \left\{ \frac{\tilde{\Lambda}_{t+1}^{cn}}{\tilde{\Lambda}_t^{cn}} \frac{1}{\Gamma_{t+1}^{x,kb}} \left(\tilde{R}_{t+1}^{cd} + (1 - \delta^{cd}) \tilde{Q}_{t+1}^{cd} \right) \right\} \quad (47)$$

$$\tilde{P}_t^{kb} = \tilde{Q}_t^{cd} \left[A_t^{cd} - 100\chi^{cd} \left(\frac{\tilde{E}_t^{cd} - \tilde{E}_{t-1}^{cd}}{\tilde{K}_t^{cd}} \Gamma_t^{x,kb} \right) \right] \quad (48)$$

$$+ \beta E_t \left\{ \frac{\tilde{\Lambda}_{t+1}^{cn}}{\tilde{\Lambda}_t^{cn}} \tilde{Q}_{t+1}^{cd} 100\chi^{cd} \left(\frac{\tilde{E}_{t+1}^{cd} - \tilde{E}_t^{cd}}{\tilde{K}_{t+1}^{cd}} \Gamma_{t+1}^{x,kb} \right) \right\}$$

$$\tilde{K}_{t+1}^{cd} = (1 - \delta^{cd}) \frac{\tilde{K}_t^{cd}}{\Gamma_t^{x,kb}} + A_t^{cd} \tilde{E}_t^{cd} - \frac{100\chi^{cd}}{2} \left(\frac{\tilde{E}_t^{cd} - \tilde{E}_{t-1}^{cd}}{\tilde{K}_t^{cd}} \Gamma_t^{x,kb} \right)^2 \frac{\tilde{K}_t^{cd}}{\Gamma_t^{x,kb}} \quad (49)$$

and

$$\tilde{Q}_t^r = \beta E_t \left\{ \frac{\tilde{\Lambda}_{t+1}^{cn}}{\tilde{\Lambda}_t^{cn}} \frac{1}{\Gamma_{t+1}^{x,cbi}} \left(\tilde{R}_{t+1}^r + (1 - \delta^r) \tilde{Q}_{t+1}^r \right) \right\} \quad (50)$$

$$\tilde{P}_t^{cbi} = \tilde{Q}_t^r \left[A_t^r - 100\chi^r \left(\frac{\tilde{E}_t^r - \tilde{E}_{t-1}^r}{\tilde{K}_t^r} \Gamma_t^{x,cbi} \right) \right] \quad (51)$$

$$+ \beta E_t \left\{ \frac{\tilde{\Lambda}_{t+1}^{cn}}{\tilde{\Lambda}_t^{cn}} \tilde{Q}_{t+1}^r 100\chi^r \left(\frac{\tilde{E}_{t+1}^r - \tilde{E}_t^r}{\tilde{K}_{t+1}^r} \Gamma_{t+1}^{x,cbi} \right) \right\}$$

$$\tilde{K}_{t+1}^r = (1 - \delta^r) \frac{\tilde{K}_t^r}{\Gamma_t^{x,cbi}} + A_t^r \tilde{E}_t^r - \frac{100\chi^r}{2} \left(\frac{\tilde{E}_t^r - \tilde{E}_{t-1}^r}{\tilde{K}_t^r} \Gamma_t^{x,cbi} \right)^2 \frac{\tilde{K}_t^r}{\Gamma_t^{x,cbi}} \quad (52)$$

The variable definitions are the same as for non-residential capital (nr) and the capital type is denoted by cd for consumer durable capital and r for residential capital.

A representative household derives utility from consumer non-durable goods and non-housing services, \tilde{E}_t^{cnn} , the flow of services from consumer-durable capital, \tilde{K}_t^{cd} , the flow of services from residential capital, \tilde{K}_t^r and leisure implicitly defined by hours worked in the two sectors, $L_t^{cbi} + L_t^{kb}$. Utility is influenced by a habit stock of each component scaled by the parameters h^{cnn} , h^{cd} and h^r . There are stochastic preference shocks to the different components denoted by Ξ_t^{cnn} , Ξ_t^{cd} , Ξ_t^r and Ξ_t^l . Households maximize utility and are monopolistic suppliers of labor. The household's budget constraint incorporates wage income, capital income, expenditure on consumption, rental payments on durable capital and residential capital, wage setting adjustment costs (depend on the parameter χ^w and the lagged and steady-state wage inflation rate) and costs in altering the composition of labor supply. Utility maximization and wage setting are constrained by the household's budget and the demand curve for the household's differentiated labor. The household's first order conditions are given by:

$$\tilde{\Lambda}_t^{cnn} = \beta R_t E_t \left\{ \tilde{\Lambda}_t^{cnn} \frac{1}{\Pi_{t+1}^{p,cbi} \Gamma_{t+1}^{x,cbi}} \right\} \quad (53)$$

$$\tilde{\Lambda}_t^{cnn} = \tilde{\Lambda}_t^{cd} \frac{1}{\tilde{R}_t^{cd}} \quad (54)$$

$$\tilde{\Lambda}_t^{cnn} = \tilde{\Lambda}_t^r \frac{1}{\tilde{R}_t^r} \quad (55)$$

$$\tilde{\Lambda}_t^{cnn} = \zeta^{cnn} \frac{\Xi_t^{cnn}}{\tilde{E}_t^{cnn} - (h^{cnn}/\Gamma_t^{x,cbi})\tilde{E}_{t-1}^{cnn}} - \beta \zeta^{cnn} E_t \left\{ \frac{(h^{cnn}/\Gamma_{t+1}^{x,cbi})\Xi_{t+1}^{cnn}}{\tilde{E}_t^{cnn} - (h^{cnn}/\Gamma_{t+1}^{x,cbi})\tilde{E}_t^{cnn}} \right\} \quad (56)$$

$$\frac{\tilde{\Lambda}_t^{cd}}{\Gamma_t^{x,kb}} = \zeta^{cd} \frac{\Xi_t^{cd}}{\tilde{K}_t^{cd} - (h^{cd}/\Gamma_{t-1}^{x,kb})\tilde{K}_{t-1}^{cd}} - \beta \zeta^{cd} E_t \left\{ \frac{(h^{cd}/\Gamma_{t+1}^{x,kb})\Xi_{t+1}^{cd}}{\tilde{K}_{t+1}^{cd} - (h^{cd}/\Gamma_{t+1}^{x,kb})\tilde{K}_t^{cd}} \right\} \quad (57)$$

$$\frac{\tilde{\Lambda}_t^r}{\Gamma_t^{x,cbi}} = \zeta^r \frac{\Xi_t^r}{\tilde{K}_t^r - (h^r/\Gamma_{t-1}^{x,cbi})\tilde{K}_{t-1}^r} - \beta \zeta^r E_t \left\{ \frac{(h^r/\Gamma_{t+1}^{x,cbi})\Xi_{t+1}^r}{\tilde{K}_{t+1}^r - (h^r/\Gamma_{t+1}^{x,cbi})\tilde{K}_t^r} \right\}, \quad (58)$$

where ζ^{cnn} , ζ^{cd} , ζ^r and ζ^l are scale parameters that tie down the ratios between the household's consumption components. $\tilde{\Lambda}_t^{cnn}$, $\tilde{\Lambda}_t^{cd}$ and $\tilde{\Lambda}_t^r$ denote marginal utility of the different goods and R_t denotes the nominal interest rate.

The household's labor-supply decisions imply the following wage Phillips curves:

$$\begin{aligned}
\Theta_t^l & \frac{\Lambda_t^{l,cbi}}{\widetilde{\Lambda}_t^{cnn}} L_t^{cbi} & (59) \\
& = (\Theta_t^l - 1) \widetilde{W}_t^{cbi} L_t^{cbi} \\
& - \Theta_t^l 100\chi^l \left(\frac{L_*^{cbi}}{L_*^{cbi} + L_*^{kb}} \widetilde{W}_t^{cbi} + \frac{L_*^{kb}}{L_*^{cbi} + L_*^{kb}} \widetilde{W}_t^{kb} \right) \left(\frac{L_t^{cbi}}{L_t^{kb}} - \eta^l \frac{L_{t-1}^{cbi}}{L_{t-1}^{kb}} - (1 - \eta^l) \frac{L_*^{cbi}}{L_*^{kb}} \right) \\
& + 100\chi^\omega \left(\Pi_t^{\omega,cbi} - \eta^\omega \Pi_{t-1}^{\omega,cbi} - (1 - \eta^\omega) \Pi_*^{\omega,cbi} \right) \Pi_t^{\omega,cbi} \widetilde{W}_t^{cbi} L_t^{cbi} \\
& - \beta E_t \left\{ \frac{\widetilde{\Lambda}_{t+1}^{cnn}}{\widetilde{\Lambda}_t^{cnn}} 100\chi^\omega \left(\Pi_{t+1}^{\omega,cbi} - \eta^\omega \Pi_t^{\omega,cbi} - (1 - \eta^\omega) \Pi_*^{\omega,cbi} \right) \Pi_{t+1}^{\omega,cbi} \widetilde{W}_{t+1}^{cbi} L_{t+1}^{cbi} \right\}
\end{aligned}$$

and

$$\begin{aligned}
\Theta_t^l & \frac{\Lambda_t^{l,kb}}{\widetilde{\Lambda}_t^{cnn}} L_t^{kb} & (60) \\
& = (\Theta_t^l - 1) \widetilde{W}_t^{kb} L_t^{kb} \\
& + \Theta_t^l 100\chi^l \left(\frac{L_*^{cbi}}{L_*^{cbi} + L_*^{kb}} \widetilde{W}_t^{cbi} + \frac{L_*^{kb}}{L_*^{cbi} + L_*^{kb}} \widetilde{W}_t^{kb} \right) \left(\frac{L_t^{cbi}}{L_t^{kb}} - \eta^l \frac{L_{t-1}^{cbi}}{L_{t-1}^{kb}} - (1 - \eta^l) \frac{L_*^{cbi}}{L_*^{kb}} \right) \\
& + 100\chi^\omega \left(\Pi_t^{\omega,kb} - \eta^\omega \Pi_{t-1}^{\omega,kb} - (1 - \eta^\omega) \Pi_*^{\omega,kb} \right) \Pi_t^{\omega,kb} \widetilde{W}_t^{kb} L_t^{kb} \\
& - \beta E_t \left\{ \frac{\widetilde{\Lambda}_{t+1}^{cnn}}{\widetilde{\Lambda}_t^{cnn}} 100\chi^\omega \left(\Pi_{t+1}^{\omega,kb} - \eta^\omega \Pi_t^{\omega,kb} - (1 - \eta^\omega) \Pi_*^{\omega,kb} \right) \Pi_{t+1}^{\omega,kb} \widetilde{W}_{t+1}^{kb} L_{t+1}^{kb} \right\}.
\end{aligned}$$

Θ_t^l denotes the elasticity of substitution between the differentiated labor inputs into production, $\Lambda_t^{l,s}$ denotes the marginal disutility of supplying labor in the two sectors, \widetilde{W}_t^s denotes the nominal wage rates and $\Pi_t^{\omega,s}$ denotes the wage inflation rates. The parameter χ^l reflects the size of adjustment costs of altering the labor supply and χ^ω the size of adjustment costs in re-setting wages. η^l determines the importance of the lagged sectoral mix of labor relative to its steady state value in the labor composition adjustment costs. η^ω determines the importance of the lagged wage inflation rate relative to its steady state value in the wage adjustment cost function.

Additionally, there are market clearing conditions and some definitional equations, for example, regarding GDP growth H_t^{gdp} and GDP deflator inflation $\Pi_t^{p,gdp}$. Finally the model is closed with a monetary policy reaction function. The nominal interest rate R_t is adjusted gradually to the central bank's target interest rate \bar{R}_t :

$$R_t = (R_{t-1})^{\phi^r} (\bar{R}_t)^{(1-\phi^r)} \exp[\varepsilon_t^r] \quad (61)$$

$$\begin{aligned}
\bar{R}_t & = \left(\Pi_t^{p,gdp} / \Pi_*^{p,gdp} \right)^{\phi^{\pi,gdp}} \left(\Delta \Pi_t^{p,gdp} \right)^{\phi^{\Delta\pi,gdp}} \\
& \quad \left(H_t^{gdp} / H_*^{gdp} \right)^{\phi^{h,gdp}} \left(\Delta H_t^{gdp} \right)^{\phi^{\Delta h,gdp}} R_*.
\end{aligned} \quad (62)$$

ε_t^r is a monetary policy shock. ϕ^r , $\phi^{\pi,gdp}$, $\phi^{\Delta\pi,gdp}$, $\phi^{h,gdp}$ and $\phi^{\Delta h,gdp}$ denote policy response parameters and R_* the steady state interest rate.

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.

All models are estimated using quarterly real-time data for real output, the output deflator and the effective federal funds rate. For the Christiano-Eichenbaum-Evans/Smets-Wouters model we use in addition real-time data for consumption, investment, hours and wages. The estimation of the model Edge et al. (2007) additionally requires data for consumption of non-durable goods and services, consumption of durable goods, residential investment, nonresidential investment, hours, wages, inflation for consumer nondurable goods and services and inflation for consumer durable goods. All time series are obtained from the Federal Reserve Bank of St. Louis' Alfred database except for hours and wages. For the 1980s and 1990s recessions we use data on aggregate weekly hours and employee compensation per hour from Faust and Wright (2009). For the 2001 and 2009 recessions we use the average weekly hours and the hourly compensation time series as in Smets and Wouters (2007) which we obtain from the Alfred database.

Consumption, investment and wages are expressed in real terms through division with the output deflator. Inflation is computed as the first difference of the log output deflator. The interest rate is expressed on a quarterly basis. Output, consumption and investment are expressed per capita by division with the civilian noninstitutional population over 16. For the 1980s and 1990s we obtain annual realtime population data from the Statistical Abstract of the United States.¹⁷ We assume a constant population growth rate within one year to construct quarterly data. For the 2001 and 2009 recessions quarterly real-time population data is available from the Alfred database.

For the 1980s and 1990s recessions we compute hours per capita by dividing aggregate hours with civilian employment (16 years and older). Realtime employment data is obtained from the Alfred database. The hours per capita series is also influenced by low frequency movements in government employment, schooling and the aging of the population that cannot be captured by the macroeconomic models. Thus, we follow Francis and Ramey (1995) and remove these trends by computing deviations of the hours per capita series using the HP filter with a weight of 16000 (compared to the standard weight of 1600 used for business-cycle frequency de-trending). The real-time character of

¹⁷Scanned documents are available as .pdf files on <http://www.census.gov/prod/www/abs/statab.html>

the data is not affected by this procedure. For the 2001 and 2009 recessions average weekly hours are multiplied with the civilian employment (16 years and older) as in Smets and Wouters (2007) to take into account the limited coverage of the nonfarm business sector compared to GDP. Finally, this hours series is expressed per capita by dividing with the population over 16.

Output, consumption, investment, wages and hours are expressed in 100 times the logarithm. Growth rates are computed as the first difference of output, consumption, investment and wages. For the FRB/EDO model we use nominal time series except for output. Inflation of nondurables and services prices and durable consumer goods prices is computed by dividing the relevant nominal and real time series.

In the forecasting exercises, per capita output growth forecasts are converted into aggregate forecasts by assuming that the average quarterly population growth of the last two years holds in the future. All data and forecasts of output growth and inflation coincide with the definition of official annualized quarterly series as we remove rounding errors of the log expressions used for the estimation of the models.