Forecasting With DSGE Models

Edward Herbst
Federal Reserve Board*

August 4, 2023

Abstract

In this chapter we revisit the forecasting performance of dynamic stochastic general equilibrium (DSGE) models. We find, consistent with prior literature, that the forecasts from a benchmark DSGE model are competitive with other forecasts. We place particular emphasis on evaluating the predictive power of a so-called behavioral New Keynesian model. For this model, we discern substantial time variation in forecasting ability, which may point towards explicit incorporation of time-varying behavioral components into these kinds of models. We also examine density forecasts from these models. Finally, we discuss the role and challenges of the zero lower bound and the COVID-19 crisis.

1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are a general and versatile class of multivariate macroeconomic models. An overarching theme of this class of models is that they are derived using economic theory: i.e., from the intertemporal behavior of (for example) households and firms, and the interactions between these agents. This theory then has implications for the business cycle behavior of key macroeconomic aggregates. This yields a powerful multivariate time series framework that can be used for many purposes including: inferring economically interpretable sources of fluctuations, studying the effects of policy analysis, and forecasting. Many central banks including the Sveriges Riskbank, European Central Bank, the Federal Reserve Bank of NY, and the Federal Reserve Board

*Preliminary. In preparation for the *Handbook of Macroeconomic Forecasting.* Correspondence: edward.p.herbst@frb.gov. This does not represent the opinions of the Federal Reserve Board or the Federal Reserve System.

use DSGE models as part of their policy analysis. This chapter will discuss the use of DSGE models for forecasting.

The application of DSGE models for forecasting has been a source of interest and debate for over 20 years. The Smets and Wouters (2007) model obtained comparable statistical fit to a flexibile vector autoregression, and a large literature subsequently emerged constructing and assessing DSGE model forecasts. Two notable contributions to this literature are Edge and Gürkaynak (2010) and Del Negro and Schorfheide (2013), who both evaluate (pseudo) real-time forecasts generated from variants of the Smets and Wouters (2007) model. The conclusion of these papers is that DSGE models are competitive in forecasting performance with both judgmental forecasts and predictions from other statistical models. Of course, whether one views this conclusion as 'glass half-full' or 'half-empty' will depend on their stance on the overall performance of macroeconomic forecasting in general. Nevertheless, the competitive nature of DSGE models reinforces their application in other areas of policy analysis.[1]

Since the research of Edge and Gürkaynak (2010) and Del Negro and Schorfheide (2013), much of the literature has focused on particular observables, episodes, or nuances of forecasting with DSGE models. For instance, Del Negro et al. (2016) emphasizes the time-varying performance of models with and without financial frictions. Cai et al. (2019) find that a DSGE model augmented with financial frictions exhibited comparable forecasting accuracy to professional forecasters in the wake of the Great Recesssion. Rubaszek (2021) uses a small scale DSGE model to predict movements in oil prices, and finds that it compares favorably. Carriero et al. (2019) employ a meta-analysis to compare forecasting models (DSGE, VAR, factor model, et cetera) and find that DSGE model competitively forecasts long-run inflation in US and UK.

In this chapter, we provide an overview of DSGE model forecasting and revisit the predictive power of DSGE models. Our study will focus on the performance of a so-called "behavioral New Keynesian" DSGE model. Behavioral New Keynesian models have emerged as a popular alternative to standard rational expectations models. These models generally feature some form of cognitive limitations of households and firms and generate forecasts and forecast errors that are more consistent with survey evidence (e.g., Coibion and Gorodnichenko (2015).) Many behavioral paradigms have been proposed in recent years—Gabaix (2016), Angeletos and Lian (2018), or Woodford (2018) —but their performance as statistical models has been less studied. Some exceptions to this are Gelain et al. (2019), who show that DSGE model with "hybrid expectations" improves forecasting for inflation, and Warne

[1]That said, Iversen et al. (2016) argue that even suboptimal DSGE model forecasts can be useful because of their interpretability and narrative power.

(2023), who finds more mixed evidence of adaptive vs. rational expectations in forecasting performance.

We focus on the finite horizon planning (FHP) model of Woodford (2018), wherein agents can reason perfectly only k periods into the future, and use a coarse estimate of how their choices will affect outcomes beyond that point. Gust et al. (2022) argue that this model tracks the time series of output growth, inflation, and interest rates prior to the Great Recession well. Their estimates point to a large departure from full information, rational expectations.

We go beyond Gust et al. (2022) and examine (pseudo) real time forecasts from this model. As is common when evaluating DSGE model forecasts, we also include the Smets and Wouters (2007) model, a larger (but with rational expectations) DSGE model. We compare the forecasts from these models to a well known benchmark: judgmental forecasts from the staff of the Federal Reserve Board prepared in advance of regularly scheduled policy meetings from 1994-2017.

We find, consistent with prior literature, that the forecasts from the Smets-Wouters model are competitive with the judgmental forecasts for GDP growth beyond short term forecasts. The inflation forecasts are somewhat worse than the judgmental ones, particularly at longer horizons. This discrepancy can be narrowed by, for instance, augmenting the Smets-Wouters model with an enlarged observable set, information about expectations, financial frictions, or other features, as prior papers have done. The FHP model does not perform as well as the Smets-Wouters model in forecasting output growth and inflation over the entire sample period. On the one hand, this result may not be surprising: the FHP model is a "small-scale" model which attributes all of the fluctuations in output growth, inflation, and the federal funds rate to just three structural shocks. On the other hand, this poor performance would seem to be at odds with the emerging interest in behavioral models for explaining key facts about expectations and the time series behavior of macroeconomic aggregates.

A deeper analysis indicates considerable time variation in the predictive power of the FHP model. In particular, the FHP model performs quite well—better than even the judgmental forecasts—in the 1990s and early part of the 2000s. The forecast performance deteriorates during the Great Recession, as is typical of models without financial frictions. However, particularly for inflation the FHP model's performance does not recover in the post-Great Recession period. This is in contrast to the Smets-Wouters model, whose performance improves in the post-Great Recession period. In particular, the FHP model consistently underpredicts the low and stable inflation that characterized the post-Great Recession period. When explaining the entire history of inflation (in addition to GDP growth and the federal funds), the model estimates short planning horizons for households and firms. Under these estimates, inflation is largely backward-looking, with the model translating the Great Recession and its low growth aftermath into extremely low inflation. By contrast, in models like Smets-Wouters the determination of current inflation weights expected future inflation more heavily, and thus they are able to better explain the low and stable inflation in the post-Great Recession period. This suggests that a time-varying degree of forward-lookingness may be an important feature of behavioral models.

We also examine density forecasts from the FHP model. The FHP model generates density forecasts that are not commensurate with the time series of output growth, inflation, and the federal funds rate. This is consistent with the FHP model's tendency to underpredict inflation. Some of this deficiency is due to the assumption of a constant variance of shocks. We find that allowing for time-varying volatility changes the density forecasts somewhat, but does not resolve the underprediction of inflation.

The results correspond to the mechanical evaluation of the predictions from 1994 to 2017 of FHP and Smets-Wouters model. The models ignores the zero lower bound (ZLB) constraint on nominal interest rates, while the sample period omits the COVID-19 recession, an extremely high-volatility period which has challenged statistical models (as well as judgmental forecasters). We discuss these two nonlinearities, the ZLB and COVID-19, from a forecasting perspective.

The chapter is organized as follows. Section 2 gives background on the mechanics of DSGE model forecasts. Section 3 describes the FHP model and the estimation data set. Section 4 evaluates point forecasts, while Section 5 examines density forecasts. Section 6 discuss some considerations related to the zero lower bound and COVID-19. Section 7 concludes.

2 The Statistical Representation of a DSGE Model

Our starting point will be a forecaster at time T with a set of observations $Y_{1:T} = \{y_t\}_{t=1}^{T}$. Their goal is to forecast $Y_{t+1:t+h}$ using a DSGE model. A generic DSGE model can be written in the following state space form:

$$y_t = \Psi(s_t, t; \theta) + u_t, \quad u_t \sim N\big(0, \Sigma_u(\theta)\big), \tag{1}$$

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta), \quad \epsilon_t \sim F_\epsilon(\cdot; \theta). \tag{2}$$

Equation (1) represents the *observation equation* which links a vector of (potentially unobserved) states s_t to a vector of observed variables y_t through the mapping Ψ. The model may also contain deterministic time trends since t is also an argument of Ψ. The formulation in equation (1) also allows for normally distributed measurement errors represented by the vector u_t . The evolution of the state variables s_t is governed by the *state equation* (2). The state equation is a function of the lagged state s_{t-1} and a vector of shocks ϵ_t. The shocks are assumed to be drawn from a known distribution F_ϵ. All of these relationships are determined by a vector of parameters θ. For DSGE, these parameters are conventionally described as "structural" or "deep" to distinguish them from the parameters of reduced form statistical models. The elements of the vector θ often have economic interpretability, e.g., the responsiveness of the monetary authority to inflation fluctuations, the representative household's coefficient of relative risk aversion, and so on. The formulation in equations (1) and (2) is quite general and allows for nonlinearities in the mappings Ψ and Φ and departures from normality in ε_t.[2] Given this model and the data $Y_{1:T}$, the first step for the forecaster is to recover plausible values for the parameter vector θ.

Parameter Estimation. It is popular to cast DSGE models as Bayesian models, where the parameters θ are treated as random variables with a prior distribution $p(\theta)$. Equations (1) and (2) can be used to construct a likelihood function, $p(Y_{1:T}|\theta)$, which is the probability of observing the data $Y_{1:T}$ given the parameters θ. The posterior distribution of the parameters is then given by Bayes' rule:

$$p(\theta|Y_{1:T}) = \frac{p(Y_{1:T}|\theta)p(\theta)}{p(Y_{1:T})}. \tag{3}$$

Analytical characterizations of the posterior $p(\theta|y_{1:T})$ are generally not available for two reasons. First, the likelihood function $p(Y_{1:T}|\theta)$ may not be known in closed form if the state space system is nonlinear or non-Gaussian. In this case, one may have to estimate the likelihood function using sequential Monte Carlo (SMC) methods such as particle filtering. Herbst and Schorfheide (2015) provide background information for this approach and Herbst and Schorfheide (2019) provide recent advances. Second, even if the system is linear and Gaussian—so that the Kalman filter may be used to compute the likelihood—the mapping from the parameters to the likelihood will be nonlinear and the posterior will not be available in closed form.

Instead, the posterior is approximated using simulation techniques. Two popular classes of approaches are Markov chain Monte Carlo (MCMC) and SMC. MCMC methods construct a Markov chain whose unique invariant distribution corresponds to the posterior in (3). By simulating from this chain, the forecaster obtains a set of (correlated) draws $\{\theta^i\}_{i=1}^N$. These draws can be used to approximate posterior expectations via Monte Carlo averages. For recent advances in MCMC for DSGE models (which can potentially sidestep the issue of likelihood estimation for nonlinear models) see Farkas and Tatar (2020) and Childers et al.

[2]Recent advances in heterogenous agent DSGE yield an infinite dimensional state vector.

(2022). Kase et al. (2022) propose a machine learning approach to model solution and parameter estimation. SMC methods are similar in spirit to MCMC methods, but instead of constructing a Markov chain, they construct a set of weighted draws $\{(\theta^i, w^i)\}_{i=1}^N$ which approximate the posterior. SMC methods can be an attractive alternative to the MCMC methods because they are more amenable to parallelization and can approximate irregular posterior distributions, which often arise in DSGE models. Herbst and Schorfheide (2015) provide background on SMC methods for DSGE models. Cai et al. (2020) adapt SMC methods to real-time "online" environments where the posterior is incrementally updated as new information arrives. This setting is particularly relevant for forecasting problems.

Forecasting. In the Bayesian paradigm, forecasts are constructed using the predictive distribution of future data, $p(Y_{T+1:T+h}|Y_{1:T})$. This predictive distribution is obtained by integrating out the parameters from the joint distribution of future data and parameters:

$$p(Y_{T+1:T+h}|Y_{1:T}) = \int p(Y_{T+1:T+h}|\theta, Y_{1:T})p(\theta|Y_{1:T})d\theta. \tag{4}$$

Like the posterior itself, the integral associated with the predictive distribution in equation (4) is typically intractable. Instead, the predictive distribution is approximated using simulation methods. One can use the draws $\{\theta^i\}_{i=1}^N$ from the posterior to construct a set $\{Y_{t+1:t+h}^i\}_{i=1}^N$ via the following procedure:

1. For each draw θ^i, simulate $s_T^i \sim s_T|\theta, Y_{1:T}$.
2. For $h = 1, \dots, H$, simulate $s_{T+h}^i \sim s_{T+h}|s_{T+h-1}^i, \theta^i$ using (2) and then $y_{T+h}^i \sim y_{T+h}|\theta^i, s_{T+h}^i$ using (1).

This procedure makes explicit the role of the states which are already integrated out of equation (4). With the draws $\{Y_{t+1:t+h}^i\}_{i=1}^N$ of the predictive distribution in hand, point and interval forecasts can be constructed. The selection of the objects minimizes an expected loss function. The loss function is determined by the forecaster. For point forecasting, common measures such as mean, median, and mode, correspond to minimizing quadratic loss, absolute loss and 0-1 loss respectively. For interval forecasting, the coverage probability of the predicted interval is often used as a measure.

3 A Small-Scale (Behavioral) New Keynesian Model

In this section, we introduce a small-scale New Keynesian (NK) DSGE model. The model shares many features with the canonical three equation NK model—textbook treatments can be found in Woodford (2003) and Galí (2008). Households face an intertemporal tradeoff between consumption and savings and monopolistically competitive firms set prices in a staggered fashion under Calvo pricing. Unlike the canonical NK model, the behavioural model presented here does not assume that households and firms set complete state-contingent plans over an infinite horizon. Instead, following Woodford (2018), households and firms exhibit a form of bounded rationality: each plans optimally only k periods into the future, taking the "continuation value" of their choices beyond k periods as given. Agents update their beliefs about this continuation value only slowly, using a form of adaptive learning. We refer to this model as the finite horizon planning or FHP model. A complete derivation of the model is beyond the scope of this review. Here, we limit ourselves to a brief description of the model's equilibrium conditions.

The economy is populated by a large number of households and firms. These households and firms may differ in their planning horizon; that is, the periods $k = 0, 1, \ldots$ that they consider in their decision-making. We assume the proportion of types of agents (both household and firms) associated with each k follows a geometric distribution with parameter $\rho \in [0, 1]$. When ρ is close to one, the mass of agents with long planning horizons is large relative to the mass of agents with short planning horizons. When ρ is close to zero, the mass of agents with short planning horizons is large relative to the mass of agents with long planning horizons. The aggregate Euler equation for output, y_t, is given by:

$$y_t - \overline{y}_t = \rho E_t \left[y_{t+1} - \xi_{t+1} - \bar{y}_{t+1} \right] - \sigma \left(i_t - \bar{i}_t - \rho E_t \left[\pi_{t+1} - \bar{\pi}_{t+1} \right] \right) + \xi_t. \tag{5}$$

The nominal interest rate and inflation rate are represented by i_t and π_t, respectively. The variable ξ_t is an exogenous demand shock. In addition to the planning horizon parameter ρ, equation (5) also contains σ, the intertemporal elasticity of substitution. The equation also includes "trend" variables $\bar{y}_t$, $\bar{\pi}_t$,and $\bar{i}_t$, whose values are determined as part of the adaptive learning process. When $\bar{y}_t = \bar{\pi}_t = \bar{i}_t = 0$ and $\rho = 1$, this equation collapses to the standard IS curve from the canonical NK model. When $\rho < 1$, equation (5) exhibits cognitive discounting.

Inflation dynamics are governed by the following Phillips curve:

$$\pi_t - \overline{\pi}_t = \beta \rho E_t \left[\pi_{t+1} - \overline{\pi}_{t+1} \right] + \kappa (y_t - y_t^* - \overline{y}_t). \tag{6}$$

The slope of the Phillips curve is given by the parameter κ, while β is the discount factor. The variable y_t^* is an exogenous supply shock. As with the IS curve, inflation dynamics follow the standard NK Phillips curve when $\rho = 1$ and $\bar{y}_t = \bar{\pi}_t = 0$.

The central bank is modeled with less behavioral realism as households and firms. Monetary policy is set according to a simple interest rate rule:

$$i_t - \bar{i}_t = \phi_\pi(\pi_t - \overline{\pi}_t) + \phi_y(y_t - \overline{y}_t) + i_t^*. \tag{7}$$

The parameters ϕ_π and ϕ_y are the coefficients governing the responsiveness of the nominal interest rate to inflation and output, respectively. The variable i_t^* is an exogenous monetary policy shock. As with the IS and Phillips curve, the monetary policy rule is a standard Taylor-type monetary policy rule when $\bar{y}_t = \bar{\pi}_t = \bar{i}_t = 0$. The trend $\bar{i}_t$ is related to the trends in $\bar{y}_t$ and $\bar{\pi}_t$ by:

$$\bar{i}_t = \bar{\phi}_\pi \bar{\pi}_t + \bar{\phi}_y \bar{y}_t. \tag{8}$$

Finally, we discuss the determination of the two trend variables $\bar{y}_t$ and $\bar{\pi}_t$. These variables are determined by the adaptive learning process. Households have an estimate of the continuation value, v_t which is updated according to the following rule:

$$v_t = (1-\gamma)v_{t-1} + \gamma v_{t-1}^{est}. \tag{9}$$

The variable v_t^{est} is an estimated value of the continuation value at time t, and $\gamma \in [0, 1)$ is the learning gain. The learning gain governs the speed at which agents update their beliefs about the value of v_t. When $\gamma = 0$, agents do not update their beliefs at all. When $\gamma = 1$, agents use the most recent observation of v_t to update their beliefs. Woodford (2018) shows that in equilibrium, the estimated value of the continuation value is given by:

$$v_t^{est} = y_t - \xi_t + \sigma\pi_t. \tag{10}$$

Averaging across different household types, Woodford (2018) shows that the effect of v_t on the spending of the average household is given by:

$$\bar{y}_t = \frac{-\sigma}{1-\rho}\left(\bar{\imath}_t - \rho\bar{\pi}_t\right) + v_t, \tag{11}$$

A similar logic holds for firms. A firm's beliefs regarding its value functions evolve according to:

$$v_{ft} = (1-\gamma_f)v_{ft-1} + \gamma_f v_{ft-1}^{est}, \tag{12}$$

where γ_f is the constant-gain learning parameter, v_{ft} denotes the effect of the continuation value on a firm's pricing decision at date t, and v_{ft}^{est} is a new estimate of that effect—which firms determine at the same time as they make their optimal pricing decision. This estimate satisfies:

$$v_{ft}^{est} = (1-\alpha)^{-1}\pi_t, \tag{13}$$

The parameter α is the degree of price stickiness in the economy. Aggregating across firms, v_{ft} can be related to trend inflation, $\bar{\pi}_t$. Woodford (2018) shows the relationship between these two variables satisfies:

$$\bar{\pi}_t = \frac{\kappa}{1-\beta\rho}\bar{y}_t + \frac{(1-\rho)(1-\alpha)\beta}{1-\beta\rho}v_{ft}. \tag{14}$$

Thus, these trend variables are time-varying, and reflect changes in these variables that arise from households and firms updating their beliefs about their value functions. Since these value functions govern the valuation of events beyond their planning horizons, including those that take place in the far future, movements in these variables can be thought of as capturing changes in the beliefs of household and firm about longer-run economic developments.

The exogenous shocks ξ_t, y_t^*, and i_t^* follow independent AR(1) processes:

$$\xi_t = \rho_\xi \xi_{t-1} + \varepsilon_t^\xi, \quad \varepsilon_t^\xi \sim N\left(0, \sigma_\xi^2\right), \tag{15}$$

$$y_t^* = \rho_y y_{t-1}^* + \varepsilon_t^y, \quad \varepsilon_t^y \sim N\left(0, \sigma_y^2\right), \tag{16}$$

$$i_t^* = \rho_i i_{t-1}^* + \varepsilon_t^i, \quad \varepsilon_t^i \sim N\left(0, \sigma_i^2\right). \tag{17}$$

The specification of the exogenous processes completes the model. The model is “small” in that it only has three shocks, but the resulting dynamics of the states can be quite rich given the presence of the learning process, without additional frictions such as habit formation or price indexation. Gust et al. (2022) evaluate in-sample fit of the model and find that it outperforms a number of competing models, including the standard NK model with rational expectations, models using detrended data, and other behavioral models, over the period 1964–2007.

Solving the model. The equations (5) – (17) constitute the model's equilibrium conditions. The states are given by:

$$s_t = \left[y_t, \pi_t, i_t, \bar{y}_t, \bar{\pi}_t, \bar{i}_t, v_t, v_t^{est}, v_{ft}, v_{ft}^{est}, \xi_t, y_t^*, i_t^*, \right]'. \tag{18}$$

Given values for the parameters, this system can be solved used standard techniques for solving linear rational expectations models, e.g., Sims (2002). The solution to this system of equations is given by a VAR(1):

$$s_t = \Phi_s(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{19}$$

The matrices $\Phi_s(\theta)$ and $\Phi_\epsilon(\theta)$ are functions of the parameters of the model, θ.
Linking to observables. The observation equations for the model are:

$$\text{Per Capita Real Output Growth}_t = \mu^Q + y_t - y_{t-1}, \tag{20}$$

$$\text{GDP Deflator Inflation}_t = \pi^A + 4 \cdot \pi_t, \tag{21}$$

$$\text{Federal Funds Rate}_t = \pi^A + r^A + 4 \cdot i_t, \tag{22}$$

where π^A and r^A are parameters governing the model's steady state inflation rate and real rate, respectively. Also, μ^Q is the growth rate of output, as we view our model as one that has been detrended from an economy growing at a constant rate, μ^Q.

Our parameter vector θ is given by:

$$\theta = \left[\beta, \sigma, \alpha, \gamma, \rho, \phi_\pi, \phi_y, \rho_\xi, \rho_y, \rho_i, \sigma_\xi, \sigma_y, \sigma_i, \mu^Q, \pi^A, r^A\right]'.$$

To complete our Bayesian model, we need to specify priors for the parameters in θ. Prior elicitation is an important part of Bayesian analysis, and can effect forecasting performance. We use relatively diffuse priors, which are described in the appendix.
A reference model. In addition to the small-scale model studied here, we present results also present results for the Smets and Wouters (2007) model, a well known benchmark in the literature. This is a medium-scale DSGE model with seven observables and was the focus of Edge and Gürkaynak (2010) and Del Negro and Schorfheide (2013). Due to space constraints and the model's extensive coverage in the literature, we have included the presentation of the model's equilibrium conditions, observation equations, and prior distribution to the appendix.

3.1 A Real Time Data Set

Macroeconomic data is often revised, either due to measurement error or because of changes to definitions. When evaluating a model's forecast accuracy, it is important to use the data that was available at the time the forecast was made. Since we use the Federal Reserve Board staff judgmental forecasts as references, we construct a real time data set whose data for any given vintage corresponds to the data available the Federal Reserve Board staff at the time. The staff at the Federal Reserve Board of Governors prepare point forecasts for key macroeconomic variables in advance of every scheduled meeting of the Federal Open Market Committee. These meetings take place approximately every six weeks, resulting in eight meetings per year. The forecasts are summarized in a document referred to as the Greenbook (to 2007) or the Tealbook (from 2007 onwards). These forecasts, which we refer to for simplicity as Tealbook forecasts, are only available to the public with a lag of five years. Therefore, the forecasts from December 2017 are the most recent. We consider only the first Tealbook forecast in each quarter beginning in 1994, which yields 96 forecast origins.

To construct a real time data set with these 96 vintages, we use the ALFRED historical database maintained by the Federal Reserve Bank of St. Louis. This effectively extends the work of Edge and Gürkaynak (2010) and Del Negro and Schorfheide (2013) beyond the Great Recession. Real time data sets can also be found from the pioneering work of Croushore and Stark (2001) and more recently McCracken and Ng (2021). The appendix contains more detailed information about the construction of the real time data set. Note that in constructing our data set, for any given vintage of data we use the most current estimate of a given time series. That is, we do not model data revisions. Galvão (2017) finds benefits to estimating DSGE models using a "released-based approach" that accounts for data revisions.

Our DSGE model is estimated on *per capita* output growth. We follow Del Negro and Schorfheide (2013) and use the most recent estimate of population growth to transform these forecasts into forecasts of total output growth.

4 Point Forecasts

In this section we examine the point forecasts associated with the small-scale model and compare it to Federal Reserve Board staff forecasts. For each of these 96 vintages of data, we estimate the model described in Section 3 using the SMC algorithm in Herbst and Schorfheide (2014). For each vintage, we use the estimated model to generate the predictive distribution for output growth, inflation, and the federal funds rate with $H = 8$ being the maximum forecast horizon. We use the mean of the predictive distribution as the point forecast. To evaluate the accuracy of these point forecasts, we first compute the root mean squared forecast error (RMSE) for each forecast horizon and each variable. For a generic variable y_t RMSE at horizon h is given by:

$$RMSE_h = \sqrt{\frac{1}{T - T_0} \sum_{t=T_0}^{T} \left(\hat{y}_{t+h|t} - y_{t+h}\right)^2}, \tag{23}$$

Here, y_{t+h} is the "actual" value of the variable at time $t + h$, $\hat{y}_{t+h|t}$ is the model's forecast of y_{t+h} at time t, T_0 denotes the start of the forecast evaluation period, and T the end of the sample. Economic data is continually revised and there is disagreement in the forecasting literature about the appropriate way to compute the actual value of the variable: while the latest available data may be closer to the "true" value, earlier releases while be more conceptually similar to the forecast itself—see the discussion in Croushore (2011). For simplicity, we use the most recent vintage of the data available at time t as the actuals, but results are not substantially affected by using the "first final" release.

Figure 1 about here.

Figure 1 plots the RMSEs of the DSGE model and the Tealbook for quarter-over-quarter output growth, quarter-over-quarter inflation, and the quarterly federal funds rate against the RMSE of the Tealbook forecasts for $h = 1, \ldots, 8$. The FRB staff projections of the federal funds rate are better understood as a conditioning assumption and so we omit this forecast from the figure. For real output growth, the FRB staff (green line) exhibits a clear advantage in nowcasting, with the $h = 1$ RMSE of about 0.40 percentage points substantially lower than either the SW model (orange line) or the FHP model (blue line). By $h = 4$, the advantage has largely disappeared. The near term advantage of a judgmental forecast like the Tealbook is not new or surprising: the forecasts considered here are typically constructed in March, June, September, and December. For example, for our 2012Q2 vintage, the Tealbook was published on June 12 thus the FRB staff forecast for 2012Q2 GDP growth incorporates all of the information released up to that date (though of course it would be difficult for the staff to "react" to new information in days leading up to publication.) Still, this is a considerable information advantage for our DSGE models, which by construction use only estimates of 2012Q1 (and prior) GDP, inflation, and the federal funds rate to generate its forecast. Aligning information sets can substantially narrow the near term difference in forecast accuracy, a point emphasized by Del Negro and Schorfheide (2013) and examined later in this chapter.

The forecasts for inflation (middle panel) indicate that the Tealbook forecasts are slightly better, in a RMSE-sense, than the DSGE model forecasts over the entire sample period. This moderate difference is increasing in the forecast horizon. The small-scale FHP model is in turn slightly worse than the SW model. For the federal funds rate (the right panel), the FHP model and the SW model give essentially the same forecasts. The RMSE is increasing quickly as the horizon increases.

This basic analysis echos the conclusions of earlier work: outside of near term real GDP forecasting, DSGE models remain competitive in terms of RMSEs with the Tealbook forecasts. This is despite the fact that the DSGE models considered here are the FHP model, a model driven by only three structural shocks, and the (nearly twenty years old) SW model. Neither model sees observable information on financial markets, expectations (either short or long term), the interest rate lower bound, and many other relevant indicators for prediction.

Figure 2 about here.

The dynamics of predictive accuracy. The results in figure 1 obscure some important details about the predictive accuracy of the models. Figure 2 shows RMSEs for the sample models, only over the 1994–2007 period, roughly the first half of the sample. Over this part of the sample, the model rankings are reversed: the point forecasts from the FHP model exhibited lower RMSEs for output growth and inflation than either the SW model or the Tealbook forecast, though the FHP model forecast for interest rates is still inferior to the SW model one. Interestingly, outside of near term GDP growth and longer horizon inflation forecasts, the FHP model also exhibits lower RMSE than the Tealbook forecasts, a difficult benchmark for any econometric model let alone as one as parsimononious as the FHP model.

Figure 3 about here.

To better understand the shifts in predictive performance **??** plots the rolling RMSE estimates based on 5 year windows for $h = 1$ (top panel) and $h = 4$ (bottom panel). In the top panel, the near-term superiority of the Tealbook is evident, the Tealbook forecast (the green line) always below the FHP (blue line) and SW (orange line) models. The Great Recession is also evident: the RMSEs for all three forecasts increase substantially after 2009, with the model-based forecast exhibiting much larger increases. Subsequent to that, forecasting performance returns roughly to its pre-financial crisis performance. For near term inflation forecasting (the top right panel), all three forecasts exhibit similar RMSEs prior to the financial crisis, with the FHP model exhibiting a slight advantage. The Great Recession was associated with a deterioration in forecasting performance of all forecasts—sometimes known as the "missing disinflation"—however both the SW model and Tealbook RMSE fall in the years following, while the RMSEs for the FHP model remain elevated. The longer horizon forecasts tell a similar story. The FHP inflation forecasts again exhibit large RMSEs in the 2010s despite performing well in the prior decades. The reason for this miss is that the FHP model expected inflation to be extremely low—even negative—for most of the 2010s. The slow growth in the wake of the extremely large negative shocks dominates the model forecast, in which it sees the "trend" component of inflation $\bar{\pi}_t$ as very low over this period. In the next subsection, we examine a simple enlargement of the FHP model observables that allows for a more accurate forecast of inflation.

4.1 The Role of Additional Information

DSGE models typically use only a small set of information available to the econometrician. While this parsimony helps facilitate the estimation and interpretation of the model, it may also be associated with poor forecasts. As seen earlier in section 4, the Federal Reserve Board staff has a considerable advantage in "nowcasting" real GDP growth over the DSGE model. The staff judgmental forecast is based on a much larger set of information than the DSGE model. Because of the larger information set, the publication delay of, say, real GDP growth data has a much smaller effect on the staff's forecast than it does on the DSGE model's forecast. Indeed, in the exercises in this paper, we ignore the fact that when constructing any given vintage, at least one third of the federal funds rate of the current quarter federal funds rate data has already been realized—because the second Tealbook of the quarter is published around the end of the second month of the quarter—or that soft indicators of current spending are available.

The presence of this information disadvantage has long been known to DSGE modelers. Boivin and Giannoni (2008) estimate a DSGE model with a much larger information set via factor modeling. Del Negro and Schorfheide (2013) argue that this information disadvantage has large effects on DSGE model forecasts, while Smets et al. (2014) show that for the Eurozone, conditioning on external information affects the point forecast accuracy. More recently, Giannone et al. (2016) explicitly model external information with a particular emphasis on higher frequency information. Some examples of approaches in the spirit—and refinements thereof—can be found in Cervená and Schneider (2014), Boneva et al. (2019), and Meyer-Gohde and Shabalina (2022). Drautzburg (2023) finds that conditioning on SPF forecasts improves DSGE model forecast accuracy.

In this chapter, we focus on incorporating a noisy measure of inflation expectations and assess how it affects point forecasts of the FHP model. First, we use a *representative agent* variant of the FHP model. In this version of the model, there is only a single k planning horizon, rather than a distribution over $k = 0, 1, \ldots$. Following the empirical evidence in Gust et al. (2022), we set $k = 1$. This delivers very similar forecasts as the heterogeneous agent model in Section 3. We augment our observables with expected inflation over the next four quarters:

$$\text{Expected Inflation}_t = \pi^A + \mathbb{E}_t^k[\pi_{t+1} + \pi_{t+2} + \pi_{t+3} + \pi_{t+4}]/4 + \eta_t \tag{24}$$

The parameter π^A is the steady state inflation rate and $\mathbb{E}_t^k$ denotes the forecast of economic agents with planning horizon of length k. We follow Del Negro and Eusepi (2011) and allow for measurement error, η_t, when including inflation expectations as an observable. The measurement error follows an AR(1) process:

$$\eta_t = \rho_\eta \eta_{t-1} + \epsilon_{\eta,t}, \text{ with } \epsilon_{\eta,t} \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right).$$

Note that the expectation used here is the agent's (non-rational) expectations, $\mathbb{E}_t^k[\cdot]$. This will not align with the expectation of inflation generated by the autoregressive solution of the model (e.g., equation (19)). Thus, it's not *a priori* obvious that this addition will improve forecasting performance of the model. Gust et al. (2023) derive the expression for this expectation. They show the FHP model with inflation expectations as an additional observable fits the data from 1966 to 2007 well in sample, and that the estimated model is consistent with many stylized facts about the relationship between inflation forecasts, forecast errors, and revisions. The formulation in equation (24) embodies the "noise" representation of additional information, rather than the "news" one, where in expectations observables contain informing regarding "future" shocks. Del Negro and Schorfheide (2013) contains more information about these two paradigms.

We use the Survey of Professional Forecasters (SPF) to construct expected GDP-deflator inflation, using the survey's mean. This survey is available in the middle of each quarter, and thus is available for the second Tealbook of each quarter. For this exercise, we re-estimate the model with inflation expectations. This allows us to get estimates for the size of the measurement error parameters ρ_η and σ_η^2 which are crucial for determining the information content of this observable. Gust et al. (2023) contains more in depth discussion of this point.

Figure 4 about here.

Figure 4 again shows rolling RMSEs for the models, this time including the FHP model which incorporates inflation expectations data as a noisy measure of the subjective expectation of the (non rational) agents expectations. The additional (red) line tracks the rolling RMSEs for output growth and inflation in the FHP model with expected inflation observable. This variant of the FHP model performs better than the baseline FHP model, even for forecasts of output growth during the crisis. More directly, the RMSE for inflation, once these inflation expectation observables are added to the model, does not stay at its elevated level. This is consistent with the evidence in Gust et al. (2023) that the FHP model with inflation expectations as an observable fits the data well in sample. While adding inflation expectations to the observation set may seem like begging the question, we stress a few points. First, these data are mapped to the subjective expectation of the model's agents, which is not the same as the expectation generated by the model's autoregressive solution. Second, the measurement error in the inflation expectations observable is estimated from the data, and is not set to zero. Third, prior to the Great Recession, augmenting the model with this additional observable does not meaningfully improve forecast accuracy.

5 Density Forecasts

While we have focused on point forecasts following much of the literature, a DSGE model can typically generate an entire probabilistic forecast (see the algorithm in Section 2). Indeed, this is one of the strengths of the DSGE model approach relative to a judgmental forecast for which generating a probabilistic forecast is difficult and extremely costly. Herbst and Schorfheide (2012) and Wolters (2015) are among the papers that have evaluated these density forecasts, building on a large statistics literature. Here we focus on whether the density forecast from the FHP model is *well calibrated*, in that the realization of events is commensurate with the probability assigned to them by the predictive distribution over a long enough sample. Dawid (1984) argues for the *prequential approach* where models are evaluated against the performance of their predictive distribution.

Of course, calibration is only one component of a good density forecast. Moreover, a density forecast can be well calibrated while not being informative: a canonical example is a weather forecast which merely reports the unconditional probability of rain. In addition, one may desire a *sharp* density forecast, that is one whose inverse variance is small. A sharp predictive distribution is one that assigns high probability to a small set of outcomes. A sharp predictive distribution may be desirable because it allows a forecaster to make more precise statements about the future.

In this chapter, however, we focus on calibration, and in particular we evaluate probability integral transformations of the density forecasts. The probability integral transformation of a random variable y is the random variable $F(y)$, where F is the cumulative distribution function of y—see Dawid (1984) and Kling and Bessler (1989) for details. If y is a random variable with a continuous distribution, then $F(Y)$ is uniformly distributed on $[0, 1]$, as shown in Rosenblatt (1952). In addition, if y is a time series $Y_{1:T}$, Diebold et al. (1998) show that the PITs are independent and identically distributed (i.i.d.) as uniform variables. We use the probability integral transformation to evaluate the calibration of the FHP model's density forecasts.

Figure 5 about here.

Figure 5 shows the histogram associated with the PITs with the FHP model for $h = 1$ (top panel) and $h = 4$ (bottom panel). The histograms are constructed using 5 bins, so that a uniform histogram would be evenly sized, each containing twenty percent of the PITs. Evidently, the PITs depart from uniformity. For $h = 1$, the left-most bin for each of the three observables is empty. This means that the FHP model assigns far too much probability to low outcomes for output growth, inflation and the federal funds rate. For output growth and the federal funds rate, the right tail of the predictive distribution also sees too few realizations, with the problem particularly pronounced for the federal funds rate. Given the persistent nature of the federal funds rate, this is not surprising. Outside the oversized left tail, the PITs for inflation for $h = 1$ are roughly well calibrated. For $h = 4$, the PITs exhibit similar characteristics: the left tail of the distribution remains too large for all three variables, and most of the realizations occur in the center of the distribution. Figure 6 displays corresponding PITs for the Smets-Wouters model. The model, though larger than the FHP model, displays similar PITs, with the left tail being oversized for all three variables at both horizons.

Figure 6 about here.

5.1 Adding Stochastic Volatility

The calibration of the density forecasts from both models is poor. A key reason for this is that sample period includes periods of considerable volatility and of low volatility, i.e., the Great Moderation, see McConnell and Perez-Quiros (2000). Justiniano and Primiceri (2008) argue that explicitly modeling this kind of time-variation is important in correctly capturing business cycle dynamics. We incorporate their approach into the FHP model. We adapt the shock specification in equations (15) – (17) as follows:

$$\xi_t = \rho_\xi \xi_{t-1} + \varepsilon_t^\xi, \quad \varepsilon_t^\xi \sim N\left(0, \sigma_{\xi,t}^2\right), \tag{25}$$

$$y_t^* = \rho_y y_{t-1}^* + \varepsilon_t^y, \quad \varepsilon_t^y \sim N\left(0, \sigma_{y,t}^2\right), \tag{26}$$

$$i_t^* = \rho_i i_{t-1}^* + \varepsilon_t^i, \quad \varepsilon_t^i \sim N\left(0, \sigma_{i,t}^2\right).. \tag{27}$$

The standard deviations $\sigma_{\xi,t}$, $\sigma_{y,t}$, and $\sigma_{i,t}$ are now time-varying and evolve according to the stochastic volatility process:

$$\ln(\sigma_{\xi,t}) = \rho_{\sigma_\xi} \ln(\sigma_{\xi,t-1}) + \eta_{\xi,t}, \quad \eta_{\xi,t} \sim N\left(0, \sigma_{\eta_\xi}^2\right), \tag{28}$$

$$\ln(\sigma_{y,t}) = \rho_{\sigma_y} \ln(\sigma_{y,t-1}) + \eta_{y,t}, \quad \eta_{y,t} \sim N\left(0, \sigma_{\eta_y}^2\right), \tag{29}$$

$$\ln(\sigma_{i,t}) = \rho_{\sigma_i} \ln(\sigma_{i,t-1}) + \eta_{i,t}, \quad \eta_{i,t} \sim N\left(0, \sigma_{\eta_i}^2\right). \tag{30}$$

Diebold et al. (2017) show that this kind of specification can help improve the performance interval and density forecasts. We denote this model as the FHPsv model and estimate the model on our 96 real time vintages of data.

Figure 7 about here.

Figure 7 displays the estimated volatilities for the final vintage of data, i.e., the 12th of

December 2017. As can be seen from the figure, there is considerable time variation in the standard deviation of the shocks. The early sample is characterized by a high volatility regime, which then falls during the Great Moderation.

Table 1 about here.

We first examine the RMSE ratios associated with the FHPsv model and compare them to the FHP model. Table 1 displays the RMSE ratios of the two models for horizons $h = 1$ and $h = 4$. As can be seen, the inclusion of stochastic volatility does not leads to an improve in point forecasting of output growth forecast at both horizons—the RMSE ratio is less than one—but a deterioration for inflation and the federal funds rate.

Figure 8 about here.

Figure 8 displays the histograms of the PITs for the FHPsv model. Relative to the constant volatility case, the distribution of the PITs at $h = 1$ is somewhat more uniform. At $h = 4$, however, now the right tail of the predictive densities for inflation and the federal funds rate are too small, with the PIT histogram exhibiting a sharp spike in its largest quintile. Essentially, the model is now too confident in its prediction about low inflation over the 2010, resulting in poor density forecasts.

6 Nonlinearities: The Zero Lower Bound and COVID

The model described in Section 3 is linear. This makes it easy to estimate, evaluate, and forecast with. However, most DSGE models are inherently nonlinear. That is, the optimality conditions the make up part of the model's equilibrium conditions are sets of nonlinear equations. And these nonlinearities are ignored in the linearized model, which locally approximates the solution near the model's steady state. Both recent advances in computation and striving for greater realism have led to the development of more empirically suited nonlinear DSGE models. The presence of large shocks to the economy, such as the financial crisis and the COVID-19 pandemic, as well as the obvious nonlinearity in the zero lower bound, and concepts like occasionally binding financial constraints have led to the development of nonlinear DSGE models. That said, nonlinear DSGE models are much more difficult to estimate and evaluate, and the value of their additional realism is not always clear for forecasting purposes. For instance, Ratto and Giovannini (2017) argue that occasionally binding financial constraints improve forecasting performance during the Great Recession, but Chin (2022) compares the forecast accuracy of a linearized version of a DSGE model against the fully nonlinear model (still ignoring the zero lower bound constraint), finding no evidence of superiority of the nonlinear model in point forecasting. Aruoba et al. (2017) argue that nonlinearities outside of obvious ones like the zero lower bound may be difficult to uncover with aggregate data.

The zero lower bound. There are two general approaches to modeling the zero lower bound: using global methods and piecewise linear methods. Global methods incorporate the zero lower bound constraint into the model's nonlinear equilibrium conditions. Examples of this include Fernández-Villaverde et al. (2015), Gust et al. (2017), and Aruoba et al. (2018). These kinds of models typically are computationally challenging and highly stylized outside of the modeling of the monetary policy. Thus far, their applicability for prediction has been limited. Piecewise linear models, on the other hand, augment the linearized equilibrium conditions with a zero lower bound constraint. Examples of this technique include Guerrieri and Iacoviello (2015), Holden (2016), Boehl (2022), Jones (forthcoming). Aruoba et al. (2020) explicitly incorporate the risk associated with an addition of the constraint, a feature missing in most of these methodologies.

The piecewise linear approach, because of its computational advantages, has seen far more use than the global solution-based one. That said, Hirose and Inoue (2016) and Atkinson et al. (2020) have little difference in parameter estimates for models with and without the zero lower bound. In practice, models estimated omitted the zero lower bound, but with forecasts conditional on the zero lower bound constraint—by adding, say, information on the expected path of the federal funds rate—tend to generate similar predictions as piecewise linear models. That said, this area remains an important topic of research—even for reduced form models, see Mavroeidis (2022). The zero lower bound is likely to be a feature of the macroeconomy for the foreseeable future. One challenge for the FHP model in particular is with short planning horizons, the FHP model is less affected by news about the future rates. While this ameliorates some undesirable features found in rational expectations models (the forward guidance puzzle), it also means that the FHP model generates similar forecasts with and without the constraint.

COVID-19. The COVID-19 pandemic was associated with a period of extremely high macroeconomic volatility as well as large monetary and fiscal shocks. The initial shock was arguably outside the scope of most DSGE models designed to capture normal business cycle fluctuations. That said, DSGE models were useful tools for generating predictions and policy counterfactuals. For example, Bodenstein et al. (2021) use a DSGE to provide plausible trajectories of the recovery from the COVID recession. Eichenbaum et al. (2022) provide an overview of epidemics embedded into New Keynesian models. Cardani et al. (2021) examine the Eurozone over this period and Lepetit and Fuentes-Albero (2022) investigate the power of monetary policy.

There is still debate on in the literature on how to incorporate the COVID period into the estimation of DSGE models. More reduced form approaches have suggested the incorporation of outliers, e.g., Antolin-Diaz et al. (2020) and Carriero et al. (2021). A difficulty here is that the DSGE model shocks all have a structural interpretation, so that assigning outliers among them is nontrivial. An additional approach would be to treat the observations in 2020Q2 and 2020Q3 as missing data, and to ignore the period when evaluating forecasts.

7 Conclusion

This chapter has focused on DSGE model forecasts, with particular emphasis on a behavioral New Keynesian model, as DSGE moves more towards behavioral realism. The forecasting performance of this model, however, is uneven, with extremely good performance in the first part of our evaluation sample, and poor performance afterwards. Incorporating more data reverses this deterioration, however, indicating that departures from rational expectations can lead to satisfactory forecasting performance.

Above all, however, the time variation in predictive power, coupled with the changing dynamics in macroeconomic data found in, for instance, the zero lower bound or COVID-19 recession, shows that model dynamic changes will be important for macroeconomic forecasting. This could take the form of explicit time-varying models or time-varying combinations of forecasts from multiple models.

References

Angeletos, G.-M. and C. Lian (2018): "Forward Guidance without Common Knowledge," *American Economic Review*, 108, 2477 2512.

Antolin-Diaz, J., T. Drechsel, and I. Petrella (2020): "Advances in Nowcasting Economic Activity: Secular Trends, Large Shocks and New Data," *SSRN Electronic Journal*.

Aruoba, B., P. Cuba-Borda, and F. Schorfheide (2018): "Macroeconomic Dynamics Near the ZLB: A Tale of Two Countries," *The Review of Economic Studies*, 85, 87 118.

Aruoba, S. B., P. Cuba-Borda, K. Higa-Flores, F. Schorfheide, and S. Villalvazo (2020): "Piecewise-Linear Approximations and Filtering for DSGE Models with Occasionally Binding Constraints," .

Aruoba, S. B. a., L. Bocola, and F. Schorfheide (2017): "Assessing DSGE model nonlinearities," *Journal of Economic Dynamics and Control*, 83, 34 54.

Atkinson, T., N. Throckmorton, and A. W. Richter (2020): "The zero lower bound and estimation accuracy," *Journal of Monetary Economics*, 115, 249–264.

Bodenstein, M., P. Cuba-Borda, J. Faris, and N. Goernemann (2021): "Forecasting During the COVID-19 Pandemic: A structural Analysis of Downside Risk," FEDS Notes. Washington: Board of Governors of the Federal Reserve System, accessed: February 01, 2021.

Boehl, G. (2022): "Efficient solution and computation of models with occasionally binding constraints," Tech. rep.

Boivin, J. and M. P. Giannoni (2008): "Optimal Monetary Policy in a Data-rich Enviroment," *Working Paper*.

Boneva, L., N. Fawcett, R. M. Masolo, and M. Waldron (2019): "Forecasting the UK economy: Alternative forecasting methodologies and the role of off-model information," *International Journal of Forecasting*, 35.

Cai, M., M. Del Negro, M. P. Giannoni, A. Gupta, P. Li, and E. Moszkowski (2019): "DSGE forecasts of the lost recovery," *International Journal of Forecasting*, 35, 1770–1789.

Cai, M., M. Del Negro, E. Herbst, E. Matlin, R. Sarfati, and F. Schorfheide (2020): "Online Estimation of DSGE Models," *The Econometrics Journal*.

Cardani, R., O. Croitorov, M. Giovannini, P. L. Pfeiffer, M. Ratto, and L. Vogel (2021): "The Euro Area's pandemic recession: A DSGE interpretation," Tech. Rep. 2021/10, JRC Working Papers in Economics and Finance.

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2021): "Addressing COVID-19 Outliers in BVARs with Stochastic Volatility," *Working paper (Federal Reserve Bank of Cleveland)*.

Carriero, A., A. B. Galvão, and G. Kapetanios (2019): "A comprehensive evaluation of macroeconomic forecasting methods," *International Journal of Forecasting*, 35, 1226–1239.

Červená, M. and M. Schneider (2014): "Short-term forecasting of GDP with a DSGE model augmented by monthly indicators," *International Journal of Forecasting*, 30, 498–516.

Childers, D., J. Fernández-Villaverde, J. Perla, C. Rackauckas, and P. Wu (2022): "Differentiable State-Space Models and Hamiltonian Monte Carlo Estimation," Tech. rep., National Bureau of Economic Research.

Chin, K. H. (2022): "Forecast evaluation of DSGE models: Linear and nonlinear likelihood," *Journal of Forecasting*, 41, 1099–1130.

Coibion, O. and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105, 2644 2678.

Croushore, D. (2011): "Frontiers of real-time data analysis," *Journal of economic literature*, 49, 72–100.

Croushore, D. and T. Stark (2001): "A real-time data set for macroeconomists," *Journal of Econometrics*, 105, 111–130.

Dawid, A. (1984): "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Series A*, 147, 278–292.

Del Negro, M. and S. Eusepi (2011): "Fitting observed inflation expectations," *Journal of Economic Dynamics and Control*, 35, 2105–2131, frontiers in Structural Macroeconomic Modeling.

Del Negro, M., R. B. Hasegawa, and F. Schorfheide (2016): "Dynamic prediction pools: An investigation of financial frictions and forecasting performance," *Journal of Econometrics*, 192, 391–405.

Del Negro, M. and F. Schorfheide (2013): "Chapter 2 - DSGE Model-Based Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 57–140.

Diebold, F., T. Gunther, and A. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Diebold, F. X., F. Schorfheide, and M. Shin (2017): "Real-time forecast evaluation of DSGE models with stochastic volatility," *Journal of Econometrics*, 201, 322 332.

Drautzburg, T. (2023): “A Structural Approach to Combining External and DSGE Model Forecasts,” Tech. Rep. 23-10, Working Papers, Federal Reserve Bank of Philadelphia.

Edge, R. and R. Gürkaynak (2010): “How Useful Are Estimated DSGE Model Forecasts for Central Bankers,” *Brookings Papers of Economic Activity*, 209–247.

Eichenbaum, M., S. Rebelo, and M. Trabandt (2022): “Epidemics in the New Keynesian Model,” *Journal of Economic Dynamics and Control*, 140.

Farkas, M. and B. Tatar (2020): “Bayesian estimation of DSGE models with Hamiltonian Monte Carlo,” Tech. rep., IMFS Working Paper Series.

Fernández-Villaverde, J., G. Gordon, P. Guerrón-Quintana, and J. F. Rubio-Ramírez (2015): “Nonlinear adventures at the zero lower bound,” *Journal of Economic Dynamics and Control*, 57, 182–204.

Gabaix, X. (2016): “A Behavioral New Keynesian Model,” .

Galí, J. (2008): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press, 2nd ed.

Galvão, A. B. (2017): “Data revisions and DSGE models,” *Journal of Econometrics*, 196, 215–232.

Gelain, P., N. Iskrev, K. J. Lansing, and C. Mendicino (2019): “Inflation dynamics and adaptive expectations in an estimated DSGE model,” *Journal of Macroeconomics*, 59, 258–277.

Giannone, D., F. Monti, and L. Reichlin (2016): “Exploiting the monthly data flow in structural forecasting,” *Journal of Monetary Economics*, 84, 201–215.

Guerrieri, L. and M. Iacoviello (2015): “OccBin: A toolkit for solving dynamic models with occasionally binding constraints easily,” *Journal of Monetary Economics*, 70, 22 38.

Gust, C., E. Herbst, and D. Lopez-Salido (2022): “Short-term Planning, Monetary Policy, and Macroeconomic Persistence,” *American Economics Journal: Macroeconomics*, forthcoming.

——— (2023): “Inflation Expectations Under Finite Horizon Planning,” *Mimeo, Federal Reserve Board*.

Gust, C., E. Herbst, D. López-Salido, and M. E. Smith (2017): "The Empirical Implications of the Interest-Rate Lower Bound," *American Economic Review*, 107, 1971 2006.

Herbst, E. and F. Schorfheide (2012): "Evaluating DSGE Model Forecasts of Comovements," *Journal of Econometrics*, 171, 152–166.

——— (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 29, 1073 1098.

——— (2015): *Bayesian Estimation of DSGE Models*, Princeton: Princeton University Press.

——— (2019): "Tempered particle filtering," *Journal of Econometrics*, 210, 26 44.

Hirose, Y. and A. Inoue (2016): "Interest rate lower bound and parameter bias in an estimated DSGE model," *Journal of Applied Econometrics*, 31, 630–651.

Holden, T. D. (2016): "Computation of solutions to dynamic models with occasionally binding constraint," *Mimeo*.

Iversen, J., S. Laseen, H. Lundvall, and U. Soderstrom (2016): "Real-Time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank," Tech. Rep. 142, Riksbank Research Paper Series, Sveriges Riksbank Working Paper Series.

Jones, C. (forthcoming): "Aging, Secular Stagnation and the Business Cycle," *Review of Economics and Statistics*.

Justiniano, A. and G. E. Primiceri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98, 604–41.

Kase, H., L. Melosi, and M. Rottner (2022): "Estimating HANK Models with Neural Networks," .

Kling, J. and D. Bessler (1989): "Calibration-Based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output," *Journal of Business*, 62, 447–499.

Lepetit, A. and C. Fuentes-Albero (2022): "The Limited Power of Monetary Policy in a Pandemic," *European Economic Review*, 147, 104–168.

Mavroeidis, S. (2022): "Identification at the Zero Lower Bound," *Econometrica*, 89, 2855–2885.

McConnell, M. M. and G. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?" *American Economic Review*, 90, 1464–76.

McCracken, M. W. and S. Ng (2021): "FRED-QD: A Quarterly Database for Macroeconomic Research," *Review, Federal Reserve Bank of St. Louis*, 103, 1–44.

Meyer-Gohde, A. and E. Shabalina (2022): "Estimation and forecasting using mixed-frequency DSGE models," Tech. Rep. 175, IMFS Working Paper Series.

Ratto, M. and M. Giovannini (2017): "Latent variables and real-time forecasting in DSGE models with occasionally binding constraints. Can Non-Linearity Improve our Understanding of the Great Recession," Mimeo.

Rosenblatt, M. (1952): "Remarks on a Multivariate Transformation," *Annals of Mathematical Statistics*, 23, 470–472.

Rubaszek, M. (2021): "Forecasting crude oil prices with DSGE models," *International Journal of Forecasting*, 37, 531–546.

Sims, C. A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.

Smets, F., A. Warne, and R. Wouters (2014): "Professional forecasters and real-time forecasting with a DSGE model," *International Journal of Forecasting*, 30, 981–995.

Smets, F. and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–608.

Warne, A. (2023): "DSGE model forecasting: rational expectations vs. adaptive learning," .

Wolters, M. H. (2015): "Evaluating point and density forecasts of DSGE models," *Journal of Applied Econometrics*, 30, 74–96.

Woodford, M. (2003): *Interest and Prices*, Princeton University Press.

——— (2018): *Monetary Policy Analysis When Planning Horizons Are Finite*, National Bureau of Economic Research, Inc, 1–50.

8 Figures and Tables

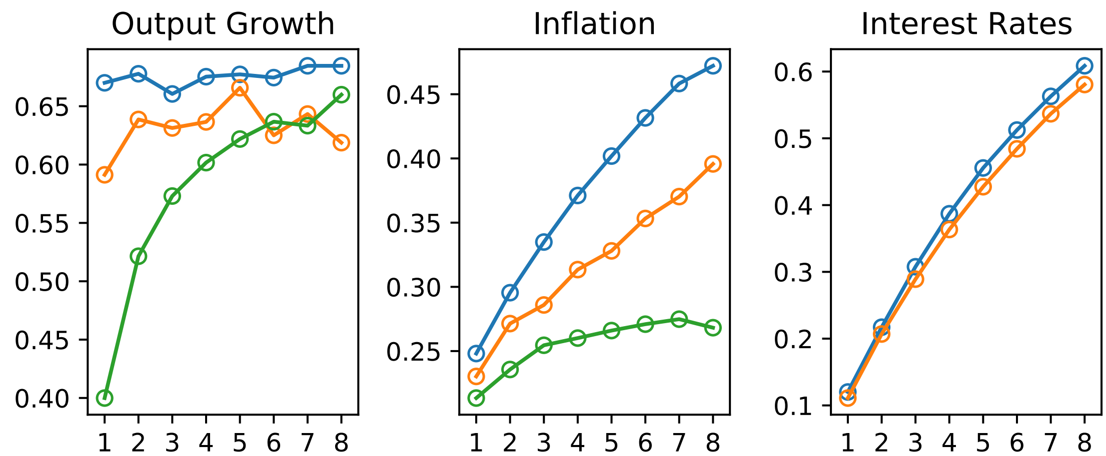


Figure 1: DSGE vs. Tealbook: 1994-2017

Notes: The figure shows RMSEs for h-step ahead forecasts for the FHP model (blue lines), the SW model (orange lines), and the FRB staff forecast (green lines) for real output growth (left panel), GDP deflator inflation (middle panel), and the federal funds rate (right panel.) All variables are expressed in quarterly terms.

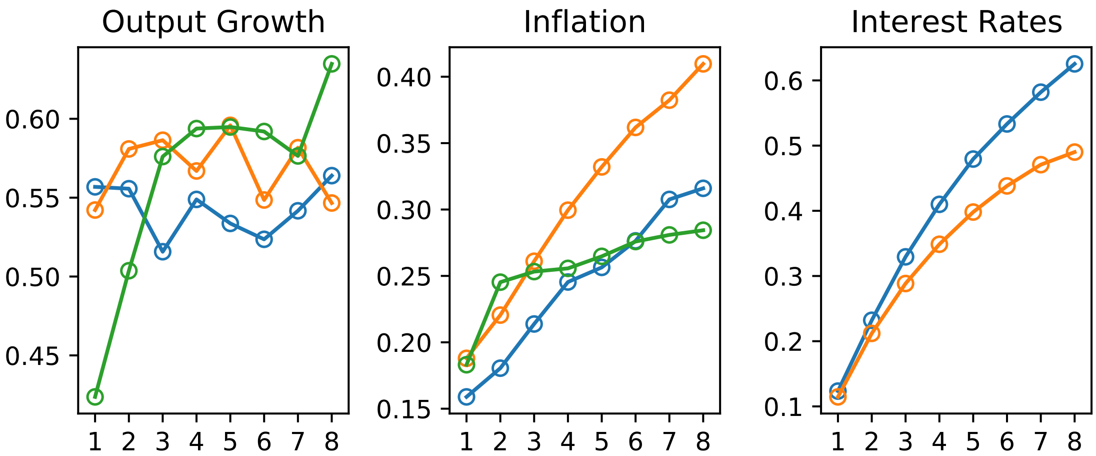
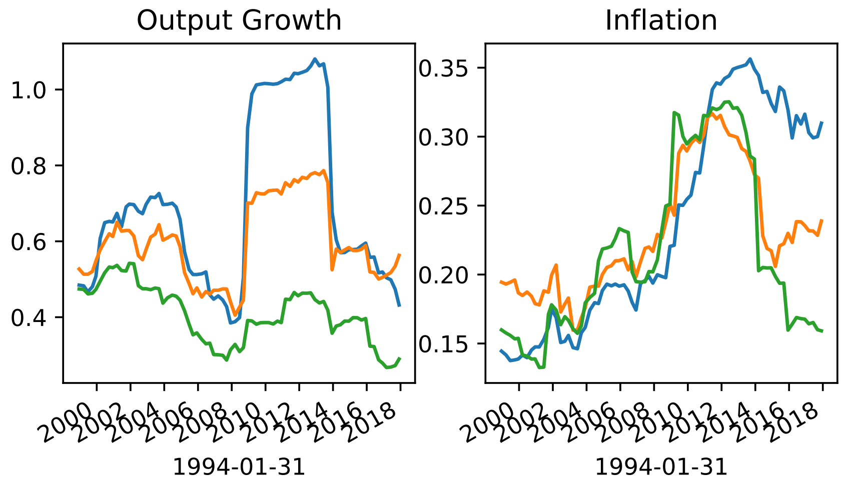


Figure 2: DSGE vs. Tealbook: 1994-2006

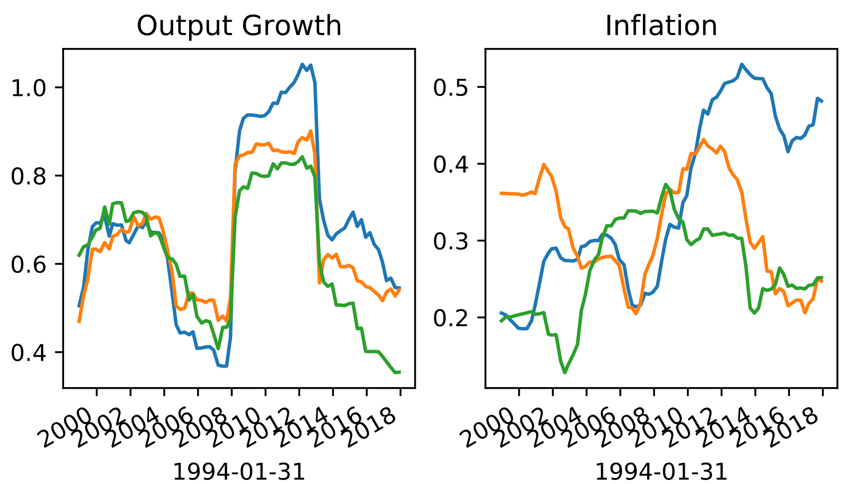
Notes: The figure shows RMSEs for h-step ahead forecasts for the FHP model (blue lines), the SW model (orange lines), and the FRB staff forecast (green lines) for real output growth (left panel), GDP deflator inflation (middle panel), and the federal funds rate (right panel.) All variables are expressed in quarterly terms.

Figure 3: 5 Year Rolling Window RMSEs

(a) $h = 1$



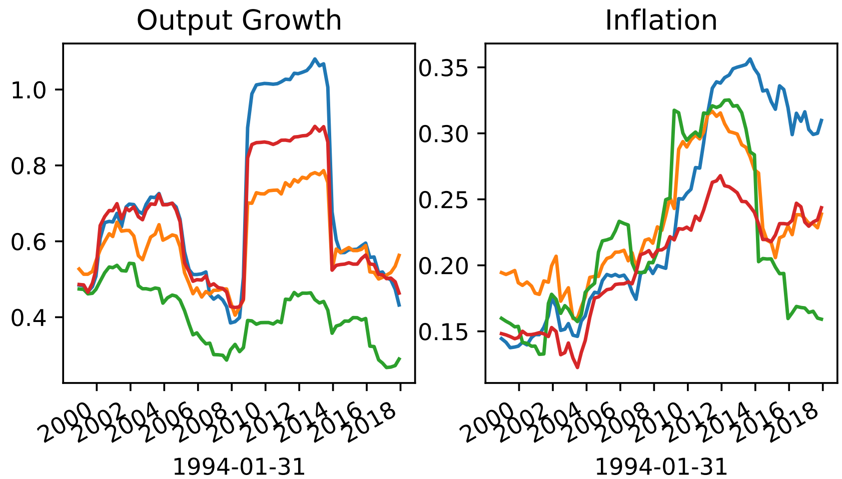
(b) $h = 4$



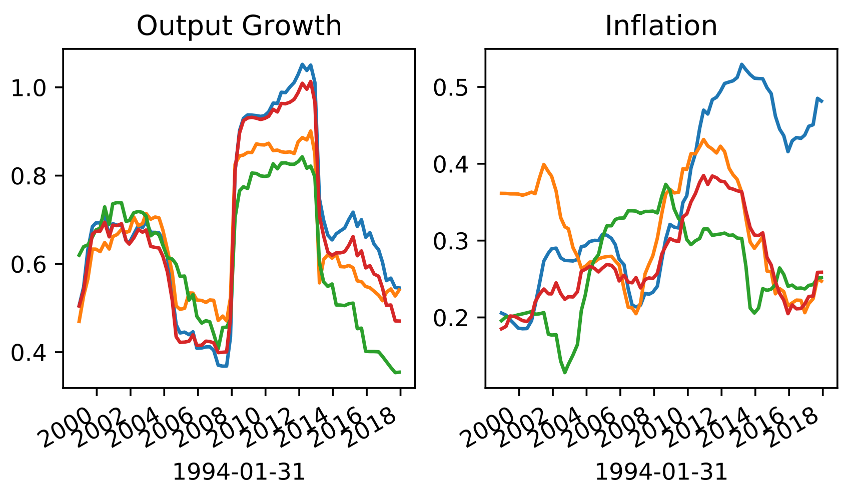
Notes: The figure shows 5 year rolling window RMSEs of 1-step ahead (top panel) and 4-step ahead (bottom panel) of forecasts for the FHP model (blue lines), the SW model (orange lines), and the FRB staff forecast (green lines) for real output growth (left panel), GDP deflator inflation (middle panel), and the federal funds rate (right panel.) All variables are expressed in quarterly terms.

Figure 4: 5 Year Rolling Window RMSEs with Additional Information

(a) $h = 1$



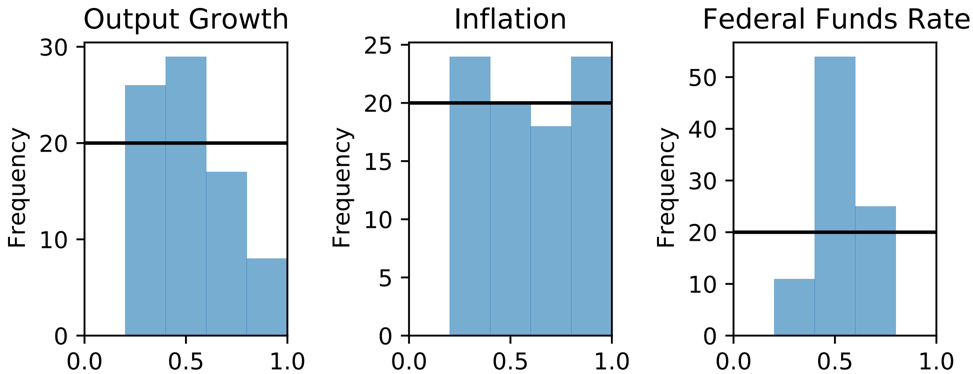
(b) $h = 4$



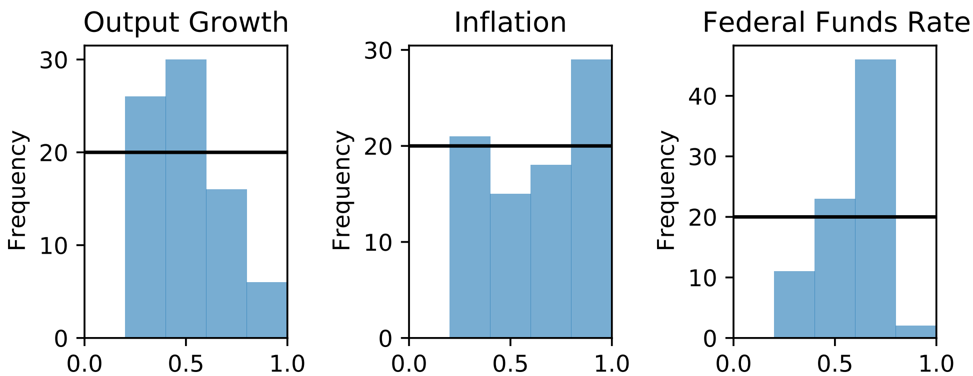
Notes: The figure shows 5 year rolling window RMSEs of 1-step ahead (top panel) and 4-step ahead (bottom panel) of forecasts for the FHP model (blue lines), the SW model (orange lines), the FRB staff forecast (green lines), and the FHP model augmented with an inflation expectations observable, for real output growth (left panel), GDP deflator inflation (middle panel), and the federal funds rate (right panel.) All variables are expressed in quarterly terms.

Figure 5: PITs for the FHP Model

(a) $h = 1$



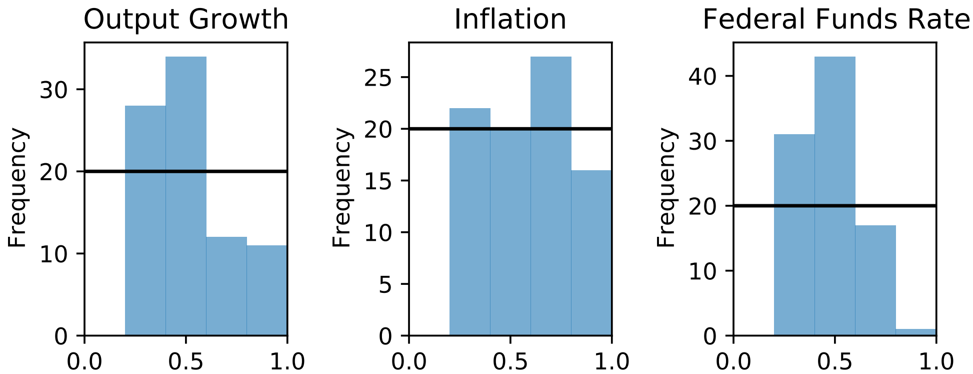
(b) $h = 4$



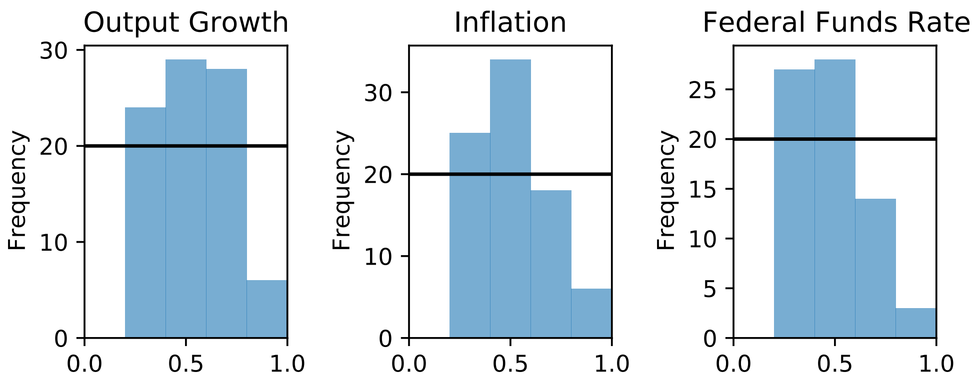
Notes: Figure show PITs for the FHP model for $h = 1$ and $h = 4$. Sample period 1994-2017.

Figure 6: PITs for the Smets-Wouters Model

(a) $h = 1$



(b) $h = 4$



Notes: Figure show PITs for the Smets-Wouters model for $h = 1$ and $h = 4$. Sample period 1994-2017.

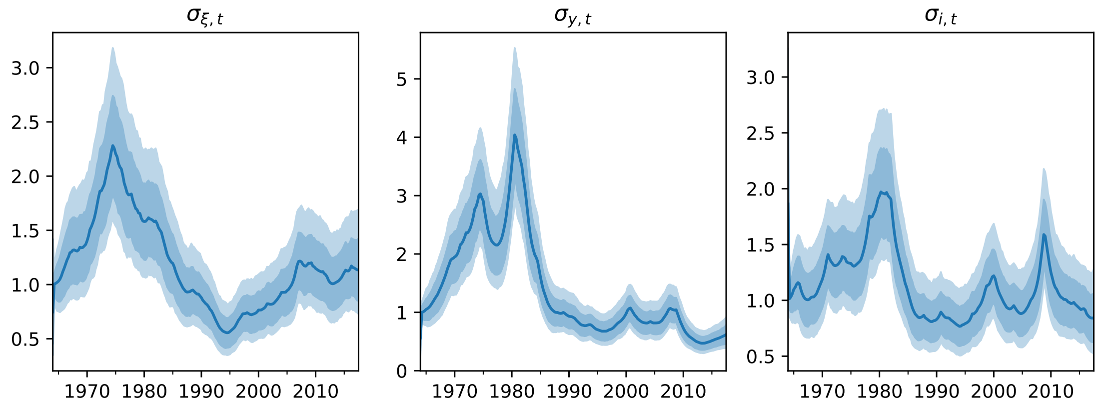


Figure 7: Estimated Stochastic Volatilities: FHP Model

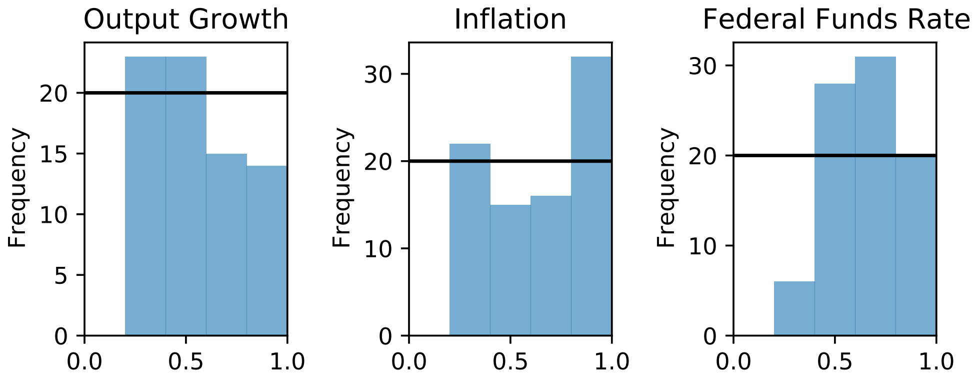
Notes: The figure shows the estimated stochastic volatilities $\sigma_{\xi,t}, \sigma_{y,t}, \sigma_{i,t}$ for the FHP model. The solid lines show the posterior mean, the light shaded region denotes the 90 percent pointwise credible interval, and the dark shaded region denotes the 68 percent pointwise credible interval.

Table 1: RMSE Ratios: FHPsv vs. FHP Model

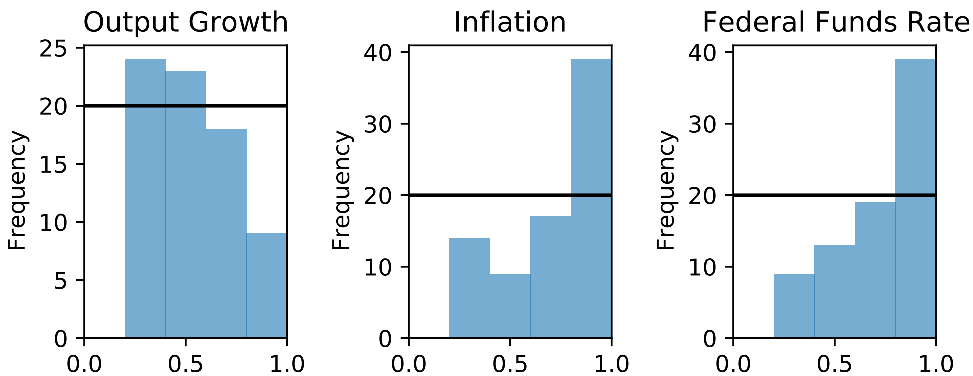
h	Output Growth	Inflation	Federal Funds Rate
1	0.85	1.02	1.07
4	0.94	1.11	1.27

Figure 8: PITs for the FHPsv Model

(a) $h = 1$



(b) $h = 4$



Appendix for "Forecasting With DSGE Models"

A Prior Distribution

Table 2: Prior Distributions

Parameter	Distribution		
	Type	Par(1)	Par(2)
r^A	Gamma	2	1
π^A	Normal	4	1
μ^Q	Normal	0.5	0.1
(ρ, γ, γ_f)	Uniform	0	1
σ	Gamma	2	0.5
κ	Gamma	0.05	0.1
ϕ_π	Gamma	1.5	0.25
ϕ_y	Gamma	0.25	0.25
$(\sigma_\xi, \sigma_{y^*}, \sigma_{i^*})$	Inv. Gamma	1	4
$(\rho_\xi, \rho_{y^*}, \rho_{i^*})$	Uniform	0	1

Table notes: Par(1) and Par(2) correspond to the mean and standard deviation of the Gamma and Normal distributions and to the upper and lower bounds of the support for the Uniform distribution. For the Inv. Gamma distribution, Par(1) and Par(2) refer to s and ν where $p(\sigma|\nu, s)$ is proportional to $\sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$.

B Details of the Data Set Construction

1. **Per Capita Real Output Growth.** Take the level of real gross domestic product, (`FRED` mnemonic "GDPC1"), call it GDP_t. Take the quarterly average of the Civilian Non-institutional Population (`FRED` mnemonic "CNP16OV" / BLS series "LNS10000000"), call it POP_t. Then,

$$\begin{aligned} &\text{Per Capita Real Output Growth} \\ &= \quad 100 \left[\ln \left(\frac{GDP_t}{POP_t} \right) - \ln \left(\frac{GDP_{t-1}}{POP_{t-1}} \right) \right]. \end{aligned}$$

2. **Annualized Inflation.** Take the GDP deflator, (`FRED` mnemonic "GDPDEF"), call it $PGDP_t$. Then,

$$\text{Annualized Inflation} = 400 \ln \left(\frac{PGDP_t}{PGDP_{t-1}} \right).$$

3. **Federal Funds Rate.** Take the effective federal funds rate (`FRED` mnemonic "FED-FUNDS"), call it FFR_t. Then,

$$\text{Federal Funds Rate} = FFR_t.$$

C The Smets-Wouters (2007) Model

The equilibrium conditions of the Smets and Wouters (2007) model take the following form:

$$
\begin{aligned}
\hat{y}_t &= c_y \hat{c}_t + i_y \hat{i}_t + z_y \hat{z}_t + \varepsilon_t^g && (31)\\
\hat{c}_t &= \frac{h/\gamma}{1+h/\gamma}\hat{c}_{t-1} + \frac{1}{1+h/\gamma}E_t\hat{c}_{t+1} + \frac{wl_c(\sigma_c-1)}{\sigma_c(1+h/\gamma)}(\hat{l}_t - E_t\hat{l}_{t+1}) && (32)\\
&\quad -\frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}(\hat{r}_t - E_t\hat{\pi}_{t+1}) - \frac{1-h/\gamma}{(1+h/\gamma)\sigma_c}\varepsilon_t^b\\
\hat{i}_t &= \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}\hat{i}_{t-1} + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}}E_t\hat{i}_{t+1} + \frac{1}{\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})}\hat{q}_t + \varepsilon_t^i && (33)\\
\hat{q}_t &= \beta(1-\delta)\gamma^{-\sigma_c}E_t\hat{q}_{t+1} - \hat{r}_t + E_t\hat{\pi}_{t+1} + (1-\beta(1-\delta)\gamma^{-\sigma_c})E_t\hat{r}_{t+1}^k - \varepsilon_t^b && (34)\\
\hat{y}_t &= \Phi(\alpha\hat{k}_t^s + (1-\alpha)\hat{l}_t + \varepsilon_t^a) && (35)\\
\hat{k}_t^s &= \hat{k}_{t-1} + \hat{z}_t && (36)\\
\hat{z}_t &= \frac{1-\psi}{\psi}\hat{r}_t^k && (37)\\
\hat{k}_t &= \frac{(1-\delta)}{\gamma}\hat{k}_{t-1} + (1-(1-\delta)/\gamma)\hat{i}_t + (1-(1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i && (38)\\
\hat{\mu}_t^p &= \alpha(\hat{k}_t^s - \hat{l}_t) - \hat{w}_t + \varepsilon_t^a && (39)\\
\hat{\pi}_t &= \frac{\beta\gamma^{(1-\sigma_c)}}{1+\iota_p\beta\gamma^{(1-\sigma_c)}}E_t\hat{\pi}_{t+1} + \frac{\iota_p}{1+\beta\gamma^{(1-\sigma_c)}}\hat{\pi}_{t-1} && (40)\\
&\quad -\frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_p)(1-\xi_p)}{(1+\iota_p\beta\gamma^{(1-\sigma_c)})(1+(\Phi-1)\varepsilon_p)\xi_p}\hat{\mu}_t^p + \varepsilon_t^p\\
\hat{r}_t^k &= \hat{l}_t + \hat{w}_t - \hat{k}_t^s && (41)\\
\hat{\mu}_t^w &= \hat{w}_t - \sigma_l\hat{l}_t - \frac{1}{1-h/\gamma}(\hat{c}_t - h/\gamma\hat{c}_{t-1}) && (42)\\
\hat{w}_t &= \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}}(E_t\hat{w}_{t+1} + E_t\hat{\pi}_{t+1}) + \frac{1}{1+\beta\gamma^{(1-\sigma_c)}}(\hat{w}_{t-1} - \iota_w\hat{\pi}_{t-1}) && (43)\\
&\quad -\frac{1+\beta\gamma^{(1-\sigma_c)}\iota_w}{1+\beta\gamma^{(1-\sigma_c)}}\hat{\pi}_t - \frac{(1-\beta\gamma^{(1-\sigma_c)}\xi_w)(1-\xi_w)}{(1+\beta\gamma^{(1-\sigma_c)})(1+(\lambda_w-1)\epsilon_w)\xi_w}\hat{\mu}_t^w + \varepsilon_t^w\\
\hat{r}_t &= \rho\hat{r}_{t-1} + (1-\rho)(r_\pi\hat{\pi}_t + r_y(\hat{y}_t - \hat{y}_t^*)) && (44)\\
&\quad + r_{\Delta y}((\hat{y}_t - \hat{y}_t^*) - (\hat{y}_{t-1} - \hat{y}_{t-1}^*)) + \varepsilon_t^r.
\end{aligned}
$$

The exogenous shocks evolve according to

$$\varepsilon_t^a = \rho_a \varepsilon_{t-1}^a + \eta_t^a \tag{45}$$
$$\varepsilon_t^b = \rho_b \varepsilon_{t-1}^b + \eta_t^b \tag{46}$$
$$\varepsilon_t^g = \rho_g \varepsilon_{t-1}^a + \rho_{ga} \eta_t^a + \eta_t^g \tag{47}$$
$$\varepsilon_t^i = \rho_i \varepsilon_{t-1}^i + \eta_t^i \tag{48}$$
$$\varepsilon_t^r = \rho_r \varepsilon_{t-1}^r + \eta_t^r \tag{49}$$
$$\varepsilon_t^p = \rho_r \varepsilon_{t-1}^p + \eta_t^p - \mu_p \eta_{t-1}^p \tag{50}$$
$$\varepsilon_t^w = \rho_w \varepsilon_{t-1}^w + \eta_t^w - \mu_w \eta_{t-1}^w. \tag{51}$$

The counterfactual no-rigidity prices and quantities evolve according to

$$\hat{y}_t^* = c_y \hat{c}_t^* + i_y \hat{i}_t^* + z_y \hat{z}_t^* + \varepsilon_t^g \tag{52}$$
$$\hat{c}_t^* = \frac{h/\gamma}{1+h/\gamma} \hat{c}_{t-1}^* + \frac{1}{1+h/\gamma} E_t \hat{c}_{t+1}^* + \frac{w l_c (\sigma_c - 1)}{\sigma_c (1+h/\gamma)} (\hat{l}_t^* - E_t \hat{l}_{t+1}^*) - \frac{1-h/\gamma}{(1+h/\gamma)\sigma_c} r_t^* - \frac{1-h/\gamma}{(1+h/\gamma)\sigma_c} \varepsilon_t^b \tag{53}$$
$$\hat{i}_t^* = \frac{1}{1+\beta\gamma^{(1-\sigma_c)}} \hat{i}_{t-1}^* + \frac{\beta\gamma^{(1-\sigma_c)}}{1+\beta\gamma^{(1-\sigma_c)}} E_t \hat{i}_{t+1}^* + \frac{1}{\varphi\gamma^2 (1+\beta\gamma^{(1-\sigma_c)})} \hat{q}_t^* + \varepsilon_t^i \tag{54}$$
$$\hat{q}_t^* = \beta(1-\delta)\gamma^{-\sigma_c} E_t \hat{q}_{t+1}^* - r_t^* + (1-\beta(1-\delta)\gamma^{-\sigma_c}) E_t r_{t+1}^{k*} - \varepsilon_t^b \tag{55}$$
$$\hat{y}_t^* = \Phi(\alpha k_t^{s*} + (1-\alpha)\hat{l}_t^* + \varepsilon_t^a) \tag{56}$$
$$\hat{k}_t^{s*} = k_{t-1}^* + z_t^* \tag{57}$$
$$\hat{z}_t^* = \frac{1-\psi}{\psi} \hat{r}_t^{k*} \tag{58}$$
$$\hat{k}_t^* = \frac{(1-\delta)}{\gamma} \hat{k}_{t-1}^* + (1-(1-\delta)/\gamma)\hat{i}_t + (1-(1-\delta)/\gamma)\varphi\gamma^2(1+\beta\gamma^{(1-\sigma_c)})\varepsilon_t^i \tag{59}$$
$$\hat{w}_t^* = \alpha(\hat{k}_t^{s*} - \hat{l}_t^*) + \varepsilon_t^a \tag{60}$$
$$\hat{r}_t^{k*} = \hat{l}_t^* + \hat{w}_t^* - \hat{k}_t^* \tag{61}$$
$$\hat{w}_t^* = \sigma_l \hat{l}_t^* + \frac{1}{1-h/\gamma} (\hat{c}_t^* + h/\gamma \hat{c}_{t-1}^*). \tag{62}$$

The steady state (ratios) that appear in the measurement equation or the log-linearized equilibrium conditions are given by

$$\gamma = \bar{\gamma}/100 + 1 \tag{63}$$

$$\pi^* = \bar{\pi}/100 + 1 \tag{64}$$

$$\bar{r} = 100(\beta^{-1}\gamma^{\sigma_c}\pi^* - 1) \tag{65}$$

$$r^k_{ss} = \gamma^{\sigma_c}/\beta - (1-\delta) \tag{66}$$

$$w_{ss} = \left(\frac{\alpha^\alpha(1-\alpha)^{(1-\alpha)}}{\Phi {r^k_{ss}}^\alpha}\right)^{\frac{1}{1-\alpha}} \tag{67}$$

$$i_k = (1-(1-\delta)/\gamma)\gamma \tag{68}$$

$$l_k = \frac{1-\alpha}{\alpha}\frac{r^k_{ss}}{w_{ss}} \tag{69}$$

$$k_y = \Phi l_k^{(\alpha-1)} \tag{70}$$

$$i_y = (\gamma - 1 + \delta)k_y \tag{71}$$

$$c_y = 1 - g_y - i_y \tag{72}$$

$$z_y = r^k_{ss}k_y \tag{73}$$

$$wl_c = \frac{1}{\lambda_w}\frac{1-\alpha}{\alpha}\frac{r^k_{ss}k_y}{c_y}. \tag{74}$$

Table 3: SW Model: Prior Distribution

Parameter	Type	Para (1)	Para (2)	Parameter	Type	Para (1)	Para (2)
φ	Normal	4.00	1.50	α	Normal	0.30	0.05
σ_c	Normal	1.50	0.37	ρ_a	Beta	0.50	0.20
h	Beta	0.70	0.10	ρ_b	Beta	0.50	0.20
ξ_w	Beta	0.50	0.10	ρ_g	Beta	0.50	0.20
σ_l	Normal	2.00	0.75	ρ_i	Beta	0.50	0.20
ξ_p	Beta	0.50	0.10	ρ_r	Beta	0.50	0.20
ι_w	Beta	0.50	0.15	ρ_p	Beta	0.50	0.20
ι_p	Beta	0.50	0.15	ρ_w	Beta	0.50	0.20
ψ	Beta	0.50	0.15	μ_p	Beta	0.50	0.20
Φ	Normal	1.25	0.12	μ_w	Beta	0.50	0.20
r_π	Normal	1.50	0.25	ρ_{ga}	Beta	0.50	0.20
ρ	Beta	0.75	0.10	σ_a	Inv. Gamma	0.10	2.00
r_y	Normal	0.12	0.05	σ_b	Inv. Gamma	0.10	2.00
$r_{\Delta y}$	Normal	0.12	0.05	σ_g	Inv. Gamma	0.10	2.00
π	Gamma	0.62	0.10	σ_i	Inv. Gamma	0.10	2.00
$100(\beta^{-1}-1)$	Gamma	0.25	0.10	σ_r	Inv. Gamma	0.10	2.00
l	Normal	0.00	2.00	σ_p	Inv. Gamma	0.10	2.00
γ	Normal	0.40	0.10	σ_w	Inv. Gamma	0.10	2.00

Notes: Para (1) and Para (2) correspond to the mean and standard deviation of the Beta, Gamma, and Normal distributions and to the upper and lower bounds of the support for the Uniform distribution. For the Inv. Gamma distribution, Para (1) and Para (2) refer to s and ν, where $p(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$.