Real-Time Forecasting Evaluation of DSGE Models with Nonlinearities

Francis X. Diebold
University of Pennsylvania

Frank Schorfheide
University of Pennsylvania

Minchul Shin
University of Pennsylvania

*** Preliminary draft ***

This Version: March 28, 2015

Abstract: Recent work has analyzed the forecasting performance of standard dynamic stochastic general equilibrium (DSGE) models, but little attention has been given to DSGE models that incorporate nonlinearities in exogenous driving processes. Against that background, we explore whether incorporating nonlinearities improves DSGE forecasts (point, interval, and density), with emphasis on stochastic volatility and regime switching. We examine real-time forecast accuracy for key macroeconomic variables including output growth, inflation, and the policy rate. We find that incorporating stochastic volatility in DSGE models of macroeconomic fundamentals markedly improves their density forecasts, just as incorporating stochastic volatility in models of financial asset returns improves their density forecasts.

Key words: Dynamic stochastic general equilibrium model, Markov switching, prediction, stochastic volatility

JEL codes: E17, E27, E37, E47

Acknowledgments: For helpful comments we thank seminar participants at the University of Pennsylvania and European University Institute, as well as Fabio Canova. For research support we thank the National Science Foundation and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia.
Contents

1 Introduction 1

2 A Benchmark DSGE Model 1

3 Model Solution and Estimation 4
   3.1 Transition ................................................. 4
      3.1.1 Constant Volatility .................................. 4
      3.1.2 Stochastic Volatility .................................. 4
      3.2 Measurement ............................................. 5
      3.3 Bayesian Estimation ..................................... 6

4 Point, Interval and Density Forecast Construction and Comparison 6
   4.1 Drawing From the Posterior Predictive Density ................. 7
   4.2 Point Forecast Construction and Comparison ................... 8
   4.3 Density Forecast Construction, Comparison and Conditional Calibration ........................................... 8
      4.3.1 Construction ............................................ 8
      4.3.2 Comparison .............................................. 8
      4.3.3 Conditional Calibration .................................. 10
   4.4 Interval Forecast Construction, Comparison and Conditional Calibration ........................................... 10
      4.4.1 Construction ............................................ 10
      4.4.2 Comparison .............................................. 10
      4.4.3 Conditional Calibration .................................. 11

5 Dataset and Procedure 11
   5.1 Real-Time Forecast Evaluation With Vintage Data ............... 11
   5.2 On the Desirability of Real-Time Analysis with Vintage Data .... 12

6 Empirical Results for Stochastic Volatility 13
   6.1 The Estimated Volatility Path ................................ 13
   6.2 DSGE Point Forecasts ....................................... 13
   6.3 DSGE Interval Forecasts ..................................... 14
   6.4 DSGE Density Forecasts ..................................... 14
   6.5 Log Predictive Density ....................................... 16

7 Empirical Results for Regime-Switching 16
   7.1 Regime Switching in Monetary Policy .......................... 16
   7.2 Regime Switching in Technological Progress .................... 16
   7.3 Empirical Results ......................................... 17

8 Concluding Remarks 17
1 Introduction

Dynamic stochastic general equilibrium (DSGE) models are now used widely for forecasting. Recently, several studies have shown that standard linearized DSGE models compete successfully with other forecasting models, including linear reduced-form time-series models such as vector autoregressions (VAR’s).\(^1\) However, little is known about the predictive importance of omitted non-linearities. Recent work by Sims and Zha (2006), Justiniano and Primiceri (2008), Bloom (2009), and Fernández-Villaverde and Rubio-Ramírez (2013) has highlighted that time-varying volatility is a key nonlinearity not only in financial data but also in macroeconomic time series. Against this background, we examine the real-time forecast accuracy (point, interval and density) of linearized DSGE models with and without stochastic volatility. We seek to determine whether and why incorporation of stochastic volatility is helpful for macroeconomic forecasting.

Several structural studies find that density forecasts from linearized standard DSGE models are not well-calibrated, but they leave open the issue of whether simple inclusion of stochastic volatility would fix the problem.\(^2\) Simultaneously, reduced-form studies such as Clark (2011) clearly indicate that inclusion of stochastic volatility in linear models (vector autoregressions) improves density forecast calibration. Our work in this paper, in contrast, is structural and yet still incorporates stochastic volatility, effectively asking questions in the tradition of Clark (2011), but in a structural environment.

We proceed as follows. In Section 2 we introduce a benchmark DSGE model, with and without stochastic volatility. In Section 3 we describe our model solution and estimation methods. In Section 4 we present aspects of point, interval and density forecasting and forecast evaluation. In Section 6 we provide empirical results on comparative real-time forecasting performance. We conclude in Section 8.

2 A Benchmark DSGE Model

Here we present our benchmark model and its equilibrium conditions. It is a small-scale New Keynesian model studied by An and Schorfheide (2007) and Herbst and Schorfheide (2012). The model economy consists of households, firms, a central bank that conducts monetary policy by setting the nominal interest rate, and a fiscal authority that determines the amount

\(^1\) See, for example, the survey of Del Negro and Schorfheide (2013).

of government consumption and finances it using lump-sum taxes. In what follows, we are summarizing the equilibrium conditions of this economy. Technology $A_t$ evolves according to

$$\log A_t = \log \gamma + \log A_{t-1} + z_t. \quad (1)$$

On average technology grows at rate $\gamma$, with exogenous fluctuations driven by $z_t$. The stochastic trend in technology induces stochastic trend in equilibrium output and consumption. The model economy has a unique deterministic steady state in terms of detrended variables $c_t = C_t/A_t$ and $y_t = Y_t/A_t$, where $C_t$ and $Y_t$ are consumption and output, respectively.

The households determine their supply of labor services to the firms and choose consumption. They receive labor and dividend income as well interest rate payments on nominal bonds. The consumption Euler equation is

$$1 = \beta \mathbb{E}_t \left[ \left( \frac{c_t}{c_{t+1}} \right)^\tau \frac{R_t}{\pi_{t+1}e^{z_t+1}} \right], \quad (2)$$

where $c_t$ is consumption, $R_t$ is the gross nominal interest rate, and $\pi_t = P_t/P_{t-1}$ is gross inflation. The parameter $\tau$ captures the relative degree of risk aversion and $\beta$ is a discount factor for the representative household.

The production sector consists of monopolistically competitive intermediate-goods producing firms and perfectly competitive final goods producers. The former hire labor from the household, produce their goods using a linear technology with productivity $A_t$, and sell their output to the final goods producers. Nominal price rigidities are introduced by assuming that the intermediate-goods producers face quadratic price adjustment costs. The final goods producers simply combine the intermediate goods. In equilibrium the inflation in the price of the final good evolves according to

$$(\pi_t - \pi) \left[ 1 - \frac{1}{2\nu} \right] \pi_t + \frac{\pi}{2\nu} = \beta \mathbb{E}_t \left[ \left( \frac{c_t}{c_{t+1}} \right)^\tau \frac{y_{t+1}}{y_t} \pi_{t+1} (\pi_{t+1} - \pi) \right] + \frac{1}{\nu \phi} \left( c_t^\tau + \nu - 1 \right), \quad (3)$$

where $\phi$ governs the price stickiness in the economy, $1/\nu$ is the elasticity of demand for each intermediate good, and $\pi$ is the steady-state inflation rate associated with the final good. Log-linarizing this equation leads to the New Keynesian Phillips curve. The larger the price adjustment costs $\phi$, the flatter the Phillips curve.

We assume that a fraction of output is used for government consumption and write

$$y_t = c_t e^{g_t}, \quad (4)$$
where $g_t$ is an exogenously evolving government spending shock. Market clearing and the aggregate resource constraint imply

$$c_t = y_t \left[ e^{-g_t} - \frac{\phi}{2} (\pi_t - \pi)^2 \right].$$

(5)

The first term in brackets captures the fraction of output that is used for government consumption and the second term captures the loss of output due to the nominal price adjustment costs. The central bank policy rule reacts to inflation and output growth,

$$R_t = R_{t-1}^{\rho_R} \left[ r \pi^* \left( \frac{\pi_t}{\pi^*} \right) \psi_1 \left( \frac{y_t}{\gamma y_{t-1}} e_t^z \right) \psi_2 \right]^{1-\rho_R} e^{m_t},$$

(6)

where $r$ is the steady state real interest rate, $\pi^*$ is the target inflation rate, and $m_t$ is a monetary policy shock.

We complete the model by specifying the exogenous shock processes,

$$m_t = \sigma_{R,t} \epsilon_{R,t}$$

(7)

$$z_t = \rho_z z_{t-1} + \sigma_{z,t} \epsilon_{z,t}$$

(8)

$$g_t = (1 - \rho_g) \log g + \rho_g g_{t-1} + \sigma_{g,t} \epsilon_{g,t}.$$  

(9)

We assume that $\epsilon_{R,t}, \epsilon_{z,t},$ and $\epsilon_{g,t}$ are orthogonal at all leads and lags, and normally distributed with zero means and unit variances.

In a constant-volatility implementation, we simply take $\sigma_{R,t} = \sigma_R$, $\sigma_{z,t} = \sigma_z$ and $\sigma_{g,t} = \sigma_g$. Incorporating stochastic volatility is similarly straightforward. Following Fernández-Villaverde and Rubio-Ramírez (2007), Justiniano and Primiceri (2008), and Fernández-Villaverde and Rubio-Ramírez (2013), we take

$$\sigma_{i,t} = \sigma_i e^{h_{i,t}}$$

$$h_{i,t} = \rho_{h} h_{i,t-1} + \nu_{i,t}$$

where $\nu_{i,t}$ and $\epsilon_{i,t}$ are orthogonal at all leads and lags, for all $i = R, z, g$ and $t = 1, \ldots, T$.  

3
3 Model Solution and Estimation

Equations (2)-(9) form a non-linear rational expectations system. The implied equilibrium law of motion can be written as

\[ s_t = \Phi(s_{t-1}, \epsilon_t; \theta) \]

where \( s_t = (y_t, c_t, \pi_t, R_t, m_t, g_t, z_t)' \) is a vector of state variables, \( \epsilon_t \) is a vector of structural shock innovations, and \( \theta \) is a vector of model parameters. In general there is no closed-form solution, so the model must be solved numerically. The most popular approach is linearization of transition and measurement equations; that is, use of first-order perturbation methods. The first-order perturbation approach yields a linear state-space system that can be analyzed with the Kalman filter.

3.1 Transition

We present transition equations with constant and stochastic volatility.

3.1.1 Constant Volatility

First-order perturbation results in a linear transition equation for the state variables,

\[ s_t = H_1(\theta)s_{t-1} + R(\theta)\epsilon_t \]

\[ \epsilon_t \sim iid N(0, Q(\theta)), \]

where \( H_1 \) is a \( ns \times ns \) matrix, \( R \) is a \( ns \times ne \) matrix and \( Q \) is a \( ne \times ne \) matrix, where \( ns \) is the number of state variables and \( ne \) is the number of structural shocks. The elements of the coefficient matrices \((H_1(\theta), R(\theta), Q(\theta))\) are non-linear functions of \( \theta \).

3.1.2 Stochastic Volatility

Linearization is inappropriate with stochastic volatility, as stochastic volatility vanishes under linearization. Instead, at least second-order approximation is required to preserve terms related to stochastic volatility, as shown by Fernández-Villaverde and Rubio-Ramírez (2007, 2013). Interestingly, however, Justiniano and Primiceri (2008) suggest a method to approximate the model solution using a partially non-linear function. The resulting law of motion is the same as that of the linearized solution, except that the variance-covariance matrix of
the structural shocks can be time-varying,

\[ s_t = H_1(\theta)s_{t-1} + R(\theta)\epsilon_t \]
\[ \epsilon_t \sim iidN(0, Q_t(\theta)). \]  
More specifically, \( Q_t(\theta) \) is a diagonal matrix with elements \( e^{2h_{i,t}} \), where each \( h_{i,t} \) has its own transition,

\[ h_{i,t} = \rho_{a_i} h_{i,t-1} + \nu_{i,t} \]
\[ \nu_{i,t} \sim iidN(0, s_{i}^2), \]  
for \( i = R, g, z \). Together with a measurement equation, equations (11) and (12) form a partially non-linear state-space representation. One of the nice features of this formulation is that the system remains linear and Gaussian, conditional on \( Q_t \).

### 3.2 Measurement

We complete the model with a set of measurement equations that connect state variables to observable variables. We consider quarter-on-quarter per capita GDP growth rates (\( YGR \)) and inflation rates (\( INF \)), and quarterly nominal interest (federal funds) rates (\( FFR \)). We measure \( INF \) and \( FFR \) as annualized percentages, and we measure \( YGR \) as a quarterly percentage. We assume that there is no measurement error. Then the measurement equation is

\[ YGR_t = 100 \log(y_t - y_{t-1} + z_t) \]
\[ INF_t = 400 \log \pi_t \]
\[ FFR_t = 400 \log R_t. \]  
In slight abuse of notation we write the measurement equation as

\[ Y_t = Z(\theta)s_t + D(\theta), \]  
where \( Y_t \) is now the \( n \times 1 \) vector of observed variables (composed of \( YGR_t, INF_t, \) and \( FFR_t \)), \( Z(\theta) \) is an \( n \times ns \) matrix, \( s_t \) is the \( ns \times 1 \) state vector, and \( D(\theta) \) is an \( n \times 1 \) vector that usually contains the steady-state values of the corresponding model state variables.
3.3 Bayesian Estimation

We perform inference and prediction using the Random Walk Metropolis (RWM) algorithm with the Kalman filter, as facilitated by the linear-Gaussian structure of our state-space system, conditional on $Q_t$. In particular, we use the Metropolis-within-Gibbs algorithm developed by Kim et al. (1998) to generate draws from the posterior distribution.

Implementing Bayesian techniques requires the specification of a prior distribution. We use priors consistent with those of Del Negro and Schorfheide (2013) for parameters that we have in common. We take the prior for the price-stickiness parameter from Herbst and Schorfheide (2012). Because $\nu$ (elasticity of demand) and $\phi$ (price stickiness) are not separately identified in the linear DSGE model, we fix $\nu$ at 0.1, and we estimate $\phi$ via $\kappa$, the slope of the Phillips curve, where $\kappa = \tau(1 - \nu)/(\nu \pi^2 \phi)$.

For the model with stochastic volatility, we specify prior distributions as:

$$\rho_{\sigma_i} \sim \mathcal{N}(0.9, 0.07), \quad s_i^2 \sim \mathcal{IG}(2, 0.05)$$

for $i = R, g, z$. We constrain the priors for the AR(1) stochastic-volatility coefficients to be in the stationary region, $\rho_{\sigma_i} \in (-1, 1)$. We set the prior mean for the variances of volatility shocks in line with the higher value used by Clark (2011), rather than the very low value used by Primiceri (2005) and Justiniano and Primiceri (2008). We summarize the complete list of prior distributions in Table 1.

4 Point, Interval and Density Forecast Construction and Comparison

In this section we discuss construction, comparison and evaluation of point, interval and density forecasts, using $p(Y_{T+1:T+H}|Y_{1:T})$, the posterior predictive density of $Y_{T+1},...,Y_{T+H}$ conditional on information available at time $T$. We begin by showing how we move from draws from the posterior parameter density, $p(\theta|Y_{1:T})$, to draws from the posterior predictive density $p(Y_{T+1:T+H}|Y_{1:T})$. We then turn to point, interval and density forecast construction and comparison.
4.1 Drawing From the Posterior Predictive Density

We generate draws from the posterior predictive density using the decomposition,

\[
p(Y_{T+1:T+H}|Y_{1:T}) = \int p(Y_{T+1:T+H}|\theta, Y_{1:T})p(\theta|Y_{1:T})d\theta
= \int \left( \int p(Y_{T+1:T+H}|S_{T+1:T+H}|s_{T+1:T+H})p(S_{T+1:T+H}|s_{T}, \theta, Y_{1:T})dS_{T+1:T+H} \right)
\times p(s_{T}|\theta, Y_{1:T})p(\theta|Y_{1:T})d(s_{T}, \theta).
\]

The decomposition shows how the predictive density reflects uncertainty about parameters, \(p(\theta|Y_{1:T})\), uncertainty about \(s_{T}\) conditional on \(Y_{1:T}\), \(p(s_{T}|\theta, Y_{1:T})\), and uncertainty about the future shocks that determine the state trajectory, \(p(S_{T+1:T+H}|s_{T}, \theta, Y_{1:T})\). Motivated by the decomposition, we generate draws from the predictive density using the following algorithm.

**Algorithm: Predictive Density Draws (Del Negro and Schorfheide (2013)):**

For \(j = 1\) to \(n_{sim}\),

1. Draw \(\theta^{(j)}\) from the posterior distribution \(p(\theta|Y_{1:T})\).

2. Using \(\theta^{(j)}\), draw \(s_{T}^{(j)}\) by using the Kalman filter to compute the mean and variance of 
   \(P(s_{T}|\theta^{(j)}, Y_{1:T})\), and generate a draw \(s_{T}^{(j)}\) from it.

3. Using \(s_{T}^{(j)}\), draw from \(p(S_{T+1:T+H}|s_{T}, \theta, Y_{1:T})\) as follows.
   
   (a) First draw a sequence of innovations \(s_{T+1:T+H}^{(j)}\).

   (b) Then, starting from \(s_{T}^{(j)}\), iterate forward the state transition equation (10), with \(\theta\) replaced by the draw \(\theta^{(j)}\), to obtain the sequence \(S_{T+1:T+H}^{(j)}\), where:

\[
s_{t}^{(j)} = H_{1}(\theta^{(j)})s_{t-1}^{(j)} + R(\theta^{(j)})\epsilon_{t}^{(j)}, \quad t = T + 1, \ldots, T + H.
\]

4. Using \(S_{T+1:T+H}^{(j)}\), draw the sequence \(Y_{T+1:T+H}^{(j)}\) using the measurement equation (14), as

\[
Y_{t}^{(j)} = D(\theta^{(j)}) + Z(\theta^{(j)})s_{t}^{(j)}, \quad t = T + 1, \ldots, T + H.
\]

The algorithm produces \(n_{sim}\) trajectories \(Y_{T+1:T+H}^{(j)}\) from the predictive distribution of \(Y_{T+1:T+H}\) given \(Y_{1:T}\).
In our subsequent empirical work, described in Section 6 below, we take 20,000 draws from the posterior distribution \( P(\theta | Y_{1:T}) \). We discard the first 5,000 draws and select every 15th draw to get \( n_{sim} = 1,000 \). For each \( j = 1, \ldots, 1,000 \), we repeat steps 3 and 4 15 times, which produces a total of 15,000 draws from the predictive distribution.

### 4.2 Point Forecast Construction and Comparison

We construct point forecasts as posterior means, which we compute by Monte Carlo averaging,

\[
\hat{y}_{T+h|T} = \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} y(j)_{T+h},
\]

The posterior mean is of course the optimal predictor under quadratic loss.

To compare the performance of point forecasts we compare root mean squared errors (RMSE’s),

\[
RMSE(i|h) = \sqrt{\frac{1}{P-h} \sum_{t=R}^{R+P-h} (y_{i,t+h} - \hat{y}_{i,t+h|t})^2},
\]

where \( R \) is the starting point of the forecast evaluation sample and \( P \) is the number of forecast origins.

### 4.3 Density Forecast Construction, Comparison and Conditional Calibration

#### 4.3.1 Construction

Density forecast construction is immediate, given the posterior predictive density. The predictive density is the density forecast.

#### 4.3.2 Comparison

We compare density forecasts using the predictive log likelihood, as in Warne et al. (2012). The predictive log likelihood is

\[
S_M(h, m) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \log p(y_{t+h} | Y_t, m), \quad h = 1, 2, \ldots, H,
\]
where $N_h$ is the number of forecasting samples, $Y_t = \{y_1, ..., y_T\}$, and $H$ is the maximum forecasting horizon. Here $M$ stands for “marginal,” as opposed to the “joint,” predictive score, defined as

$$ S_J(h, m) = \frac{1}{N_h} \sum_{t=T}^{T+N_h-1} \log p(y_{t+1}, ..., y_{t+h}|Y_t, m), \quad h = 1, 2, ..., H. $$

The marginal predictive score has a nice interpretation. To see this we decompose $p(\cdot)$ in equation (16)

$$ p(y_{t+h}|Y_t) = \frac{p(y_{t+h}, Y_t)}{p(Y_t)}, \quad h = 1, 2, ..., H. $$

Obviously the joint and marginal predictive score concepts lead to the same quantity when $h = 1$. For $h > 1$, to calculate the predictive density, we treat $y_{t+1}, ..., y_{T+h-1}$ as missing values and apply the Kalman filter.

To evaluate the predictive likelihood we proceed as follows.\(^3\) Note that

$$ p(Y_{T+h}, h_{T+1}, ..., h_{T+h-1}, \tilde{\theta}|Y_T) = p(Y_{T+h}|h_{T+1}, ..., h_{T+h-1}, \tilde{\theta}, Y_{1:T}) \times p(h_{T+1}, ..., h_{T+h-1}, \tilde{\theta}|Y_T) $$

$$ = p(Y_{T+h}|h_{T+1}, ..., h_{T+h-1}, \tilde{\theta}, Y_{1:T}) $$

$$ \times p(h_{T+1}, ..., h_{T+h-1}|\tilde{\theta}, Y_{1:T}) \times p(\tilde{\theta}|Y_{1:T}) $$

where $\tilde{\theta} = (\theta, h_{1:T})$. Hence

$$ p(Y_{T+h}|Y_{1:T}) = \int ... \int p(Y_{T+h}, h_{T+1}, ..., h_{T+h-1}, \tilde{\theta}|Y_{1:T})dh_{T+1}...dh_{T+h-1}d\tilde{\theta}. $$

This suggests the following algorithm.

**Algorithm: Predictive log likelihood:**

For $s = 1, ..., S$,

a) Draw $\tilde{\theta}^s = (\theta^s, \{h_{ts}^s\}_{t=1}^T) \sim p(\theta|Y_{1:T})$.

b) Generate $\{h_{T+1}^{s,i}, ..., h_{T+h}^{s,i}\}$ for $i = 1, ..., n_{traj}$.

c) Evaluate $\log p\left(Y_{T+h}|h_{T+1}^{s,i}, ..., h_{T+h}^{s,i}, \tilde{\theta}^s, Y_{1:T}\right)$ for $i = 1, ..., n_{traj}$.

d) Form $\log p(Y_{T+h}|Y_{1:T}) \approx \frac{1}{S} \frac{1}{n_{traj}} \sum_{s=1}^S \sum_{i=1}^{n_{traj}} \log p\left(Y_{T+h}|h_{T+1}^{s,i}, ..., h_{T+h}^{s,i}, \tilde{\theta}^s, Y_{1:T}\right)$.\(^3\)

\(^3\)We consider the time-varying volatility model. Calculation for the constant-volatility model is just a special case.
4.3.3 Conditional Calibration

The predictive log likelihood density forecast comparison approach described above invokes a relative standard; using the log predictive density, it ranks density forecasts according to assessed likelihoods of the observed realization sequence. It is also of general interest to assess density forecasts relative to a different, absolute standard, correct conditional calibration.

Following Diebold et al. (1998), we rely on the probability integral transform (PIT). The PIT of \( y_{i,T+h} \) based on the time-\( T \) predictive distribution is defined as the cumulative density of the random variable \( y_{i,T+h} \) evaluated at the true realization of \( y_{i,T+h} \),

\[
  z_{i,h,T} = \int_{-\infty}^{y_{i,T+h}} p(\tilde{y}_{i,T+h}|Y_{1:T})d\tilde{y}_{i,T+h}.
\]

We compute PITs by the Monte Carlo average of the indicator function,

\[
  z_{i,h,T} \approx \frac{1}{n_{\text{sim}}} \sum_{j=1}^{n_{\text{sim}}} I\{y^{(j)}_{i,T+h} \leq y_{i,T+h}\}.
\]

If the predictive distribution is correctly conditionally calibrated, then \( z_{i,h,T} \) should be distributed \( U(0,1) \) and be at most \( h-1 \)-dependent.

4.4 Interval Forecast Construction, Comparison and Conditional Calibration

4.4.1 Construction

Posterior interval forecast (credible region) construction is immediate, given the posterior predictive density, as the interval forecast follows directly from the predictive density. We focus on single-variable credible intervals as opposed to multi-variable credible regions. We compute the highest-density 100(1 - \( \alpha \)) percent interval forecast for a particular element \( y_{i,T+h} \) of \( y_{T+h} \) by numerically searching for the shortest connected interval that contains 100(1 - \( \alpha \)) percent of the draws \{\( y^{(j)}_{i,T+h} \)\}_{j=1}^{n_{\text{sim}}}. 

4.4.2 Comparison

We consider both interval forecast coverage and length, recognizing the tradeoff, precisely as with the bias-variance tradeoff for point forecasts.
4.4.3 Conditional Calibration

As detailed in Christoffersen (1998), if interval forecasts are correctly conditionally calibrated, then the “hit sequence” should have mean \(1 - \alpha\) and be at most \(h - 1\)-dependent, where the hit sequence is \(I_t^{(1-\alpha)} = 1\{\text{realized } y_t \text{ falls inside the interval}\}\). Note well the two-part characterization. The hit series must have the correct mean, \((1 - \alpha)\), which corresponds to correct unconditional calibration, and it must also be at most \(h - 1\)-dependent. When both hold, we have correct conditional calibration.

5 Dataset and Procedure

5.1 Real-Time Forecast Evaluation With Vintage Data

For the evaluation of point and density forecasts, we use the real-time data set constructed by Del Negro and Schorfheide (2013).\(^4\) They compare the point forecasts from the DSGE model to those from the Bluechip survey and Federal Reserve Board’s Greenbook. To make forecasts from the DSGE models comparable to the real-time forecasts made in Bluechip survey and Greenbook, they built data vintages that are aligned with the publication dates of the Blue Chip survey and the Federal Reserve Board’s Greenbook. In this study, we use the data set matched with the Blue Chip survey publication dates.

The first forecast origin considered in the forecast evaluation is (two weeks prior to) January 1992 and the last forecast origin for one-step-ahead forecasts is (two weeks prior to) April 2011. As in Del Negro and Schorfheide (2013), we use data vintages in April, July, October, and January. Hence we make use of four data vintages per year and the total number of data vintages used in this study is 78. The estimation sample starts from 1964:Q2 for all vintages. Following Del Negro and Schorfheide (2013), we compute forecast errors based on actuals that are obtained from the most recent vintage hoping that these are closer to the “true” actuals\(^5\). \(^6\)

To evaluate forecasts we recursively estimate DSGE models over the 78 vintages starting from the January 1992 vintage. That is, for the January 1992 vintage we estimate DSGE

\(^4\)The same data set used in Herbst and Schorfheide (2012).

\(^5\)Alternatively, we could have used actuals from the first “final” data release, which for output corresponds to the “Final” NIPA estimate (available roughly three months after the quarter is over). Del Negro and Schorfheide (2013) found that the general conclusions about the forecast performance of DSGE models are not affected by the choice of actuals. Similar conclusion is made in Rubaszek and Skrzypczyński (2008).

\(^6\)More details on the data set can be found in the Del Negro and Schorfheide (2013).
models with sample from 1964:Q2 to 1991:Q3 and generate forecasts for 1991:Q4 (one-
step-ahead) ~ 1993:Q2 (eight-steps-ahead). We iterate this procedure for all vintages from
January 1992 to April 2011.

5.2 On the Desirability of Real-Time Analysis with Vintage Data

From a model-selection perspective, one might ask whether a full-sample analysis with final-
revised data, as opposed to an expanding-sample analysis with real-time vintage data, would
be more informative. For our purposes in this paper the answer is clearly no, because our
interest is intrinsically centered on real-time performance, which is an expanding-sample
phenomenon involving vintage data. That is, each period we get not only a new observation,
but also an improved estimate of the entire history of all observations. Analysis based on
final-revised data, even pseudo-real-time analysis based on an expanding sample, is simply
not relevant.

Let us consider real-time vintage data issues from a more formal Bayesian viewpoint
centered on predictive likelihood in its relation to marginal likelihood. By Bayes’ theorem
the predictive likelihood is a ratio of marginal likelihoods,

\[ p(y_{t+1} | \tilde{y}_t, M_i) = \frac{p(\tilde{y}_{t+1} | M_i)}{p(\tilde{y}_t | M_i)}, \]

so that

\[ \prod_{t=1}^{T-1} p(y_{t+1} | \tilde{y}_t, M_i) = \frac{p(\tilde{y}_T | M_i)}{p(\tilde{y}_1 | M_i)}. \]

Hence one can say that Bayesian model selection based on the full-sample predictive perfor-
mance record and based on the full-sample marginal likelihood are the same.

The crucial insight is that in our context “full-sample” should not just refer to the full
sample of final-revised data, but rather the union of all samples of vintage data, so we
now introduce notation that distinguishes between the two. Let \( \tilde{y}_t^{(T)} \) be the data up to
time \( t \) viewed from the time-\( T \) vantage point (vintage \( T \)), and let \( \tilde{y}_t^{(t)} \) be the data up to
time \( t \) viewed from the time-\( t \) vantage point (vintage \( t \)). In our more refined notation, the
predictive-likelihood Bayesian model selection prescription is not \( \prod_{t=1}^{T-1} p(y_{t+1} | \tilde{y}_t^{(T)}, M_i) \), but
rather \( \prod_{t=1}^{T-1} p(y_{t+1} | \tilde{y}_t^{(t)}, M_i) \). That is precisely what we do.

\[ ^7 \text{See Diebold (2015).} \]
6 Empirical Results for Stochastic Volatility

6.1 The Estimated Volatility Path

For the DSGE model with stochastic volatility to produce density forecasts more accurate than constant volatility DSGE models, the estimated variances of structural shocks should change significantly over time. Therefore, as a starting point, we overlay and compare estimated variances for structural shocks from the constant volatility DSGE model and stochastic volatility DSGE model. Figure 1 reports estimates (posterior means) obtained with different real-time data vintages. The stochastic volatility estimates (solid line) and constant volatility estimate (dotted line) are obtained with data samples ending in 1991:Q3, 2001:Q3, and 2010:Q3 (obtained from data vintages of January 1992, January 2002, and January 2011, respectively). Overall, the estimates confirm significant time variation in volatility. In particular, volatility fell sharply in the mid-1980s with the Great Moderation. The estimates also reveal a sharp rise in volatility in recent years (for volatility on the technology shock). The general shapes of volatility are very similar across vintages, but level can differ slightly. Standard deviation estimates from the constant volatility DSGE model overstate the standard deviations for the post-Great Moderation period because the model tries to balance out the effect of the high variance before the 80s. Estimates are decreasing as we include more data from the Great Moderation, but they are still larger than the estimates from the model with stochastic volatility.

6.2 DSGE Point Forecasts

Figure 2 and Table 2 present real-time forecasts RMSEs for 1991:Q4-2011:Q1. Table 2 includes RMSEs for constant variance linearized DSGE model (benchmark) forecasts and ratios of RMSEs for stochastic volatility DSGE model. In these blocks, entries with value of less than 1 mean that a forecast is more accurate than the benchmark model. To provide a rough measure of statistical significance, Table 2 presents $p$-values for the null hypothesis that the MSE of a given model is equal to the MSE of the constant volatility linearized DSGE model, against the (one-sided) alternative that the MSE of the given model is lower. These $p$-values are obtained by comparing the tests of Diebold and Mariano (1995) against standard normal critical values.

The model with stochastic volatility tends to generate more accurate point forecasts for output growth especially in the short-run. It also has a smaller RMSE for the interest rate in
the very short horizon \((h = 1)\). However, inflation forecasts are very similar across models. In general, RMSEs of the DSGE model with stochastic volatility get closer to the constant variance DSGE model as the horizon increases.

6.3 DSGE Interval Forecasts

Table 3 reports the frequency with which real-time outcomes for output growth, inflation rate, and the federal funds rate fall inside 70% highest posterior density intervals estimated in real time with the DSGE models. Accurate intervals should produce frequencies of about 70%. A frequency of greater than (less than) 70% means that on average over a given sample, the posterior density is too wide (narrow). The table includes \(p\)-values for the null of correct coverage (empirical = nominal rate of 70%), based on \(t\)-statistics. These \(p\)-values are provided as a rough gauge of the importance of deviations from correct coverage.

As Table 3 shows, the constant variance DSGE models tend to be too wide, with actual outcomes falling inside the intervals much more frequently than the nominal 70% rate. For example, for the one-step-ahead forecast horizon, the linearized DSGE model coverage rates range from 91% to 95%. Based on the reported \(p\)-values, all of these departures from the nominal coverage rate appear to be statistically meaningful. For all cases, coverage rates are much wider than the nominal rate, meaning that the constant volatility DSGE model overestimates the uncertainty in the predictive distribution.

Adding stochastic volatility to the DSGE model improves the calibration of the interval forecasts with coverage rates closer to the nominal rate than the constant variance DSGE models. For example, at horizon one, coverage rate for output growth decreases to 77% from 95%. However, they are still somewhat higher for some cases having coverage rates higher than 80%.

6.4 DSGE Density Forecasts

Figure 3 and 4 report histograms of probability integral transformation (PIT) for horizon 1 and 4, respectively. PITs are grouped into five equally sized bins. Under a uniform distribution, each bin should contain 20% of the PITs, indicated by the solid horizontal lines in the figure. For output growth, large fraction of PITs fall into 0.4-0.8 bins, indicating that the predictive distribution is too diffuse. PITs for inflation rate and federal funds rate have a similar problem. Tails are not covered and too few fractions of PITs are in the 0-0.2 bin (left tail) for inflation rate or in the 0.8-1 bin (right tail) for federal funds rate, showing that
uncertainty is overestimated by the density forecasts. For longer horizon \( h = 4 \), histograms for federal fund rate becomes closer to the uniform distribution, but still only a small fraction of PITs fall into the 0.8-1 bin (right tail).

The inclusion of stochastic volatility to the DSGE model substantially improves the calibration of the density forecasts. The PIT histograms are more close to that of the uniform distribution. For example, at horizon 1, PITs for output growth tend to be equally distributed. More PITs are distributed on the tail for both inflation rate and federal funds rate. Although the discrepancy between histograms and the horizontal line become larger as the horizon increases, they are closer to the uniform distribution than in the constant volatility case.

For a more formal assessment, Table 6 reports various test metrics, including the standard deviations of the normalized error, along with \( p \)-values for the null that the standard deviation equals 1; the means of the normalized errors, along with \( p \)-values for the null of a zero mean; the AR(1) coefficient estimate and its \( p \)-value, obtained by a least squares regression including a constant; and the \( p \)-value of Berkowitz (2001)’s likelihood ratio test for the joint null of zero mean, unity variance, and no AR(1) serial correlation.

The tests confirm that without stochastic volatility, standard deviations are much less than 1, means are sometimes nonzero, and serial correlation can be considerable. For example, standard deviations of normalized forecast errors from the constant volatility DSGE model are at most 0.58 for output growth and federal funds rate; the corresponding \( p \)-values are all close to 0. For inflation rate, standard deviations are slightly larger but still smaller than 1. The absolute values of AR (1) coefficients range from 0.12 to 0.66, with \( p \)-values close to 0 except in the case of inflation for which the \( p \)-value is 0.47. Not surprisingly, given the results for means, standard deviations, and AR (1) coefficients, the \( p \)-values of the Berkowitz (2001) test are nearly 0.

By allowing stochastic volatility, the quality of density forecasts is improved based on the formal metrics. The standard deviations of the normalized forecast errors are all greater than 0.8 with \( p \)-values of 0.349 or greater. Means are all close to zero, whereas the means for output growth from the constant volatility DSGE model are around 0.2, with \( p \)-value smaller than 0.005. The patterns for AR (1) coefficients are similar to the constant volatility case, except for inflation rate (no AR (1) serial correlation). Due to the high AR (1) coefficient, \( p \)-values for LR tests are also nearly 0 for output growth and federal funds rate, even after adding the stochastic volatility.
6.5 Log Predictive Density

Table 7 presents the average log predictive densities (from one-step-ahead to eight-step-ahead). The model with stochastic volatility has a substantially larger value, which again confirms that the density forecasts are improved by including stochastic volatility ($h = 1$). However, as $h$ increases constant volatility model has larger log predictive density. Figure 5 shows the time-series plot of the one-step-ahead and four-step-ahead log predictive densities. For most of the time, the one-step-ahead predictive densities of the model with stochastic volatility are higher than the constant volatility DSGE model.

7 Introducing Regime Switching

Here we consider incorporating regime-switching nonlinearities.

7.1 Regime Switching in Monetary Policy

We can incorporate regime-switching nonlinearity in monetary policy by letting the target inflation rate evolve according to a two-state Markov-switching process, as in Schorfheide (2005). Specifically, we replace the central bank policy rule in equation (6) with

$$R_t = R_{t-1}^{ρR} \left[ r \left( \frac{π_t^*}{π_t} \right) ψ_1 \left( \frac{y_t}{γy_{t-1}} e_t^z \right)^{ψ_2} \right]^{1-ρR} e^{mt}, \quad (17)$$

where the target inflation rate switches between two states ($s_{p,t} = 1, 2$),

$$π_t^* = \begin{cases} π_L & \text{if } s_{p,t} = 1 \\ π_H & \text{if } s_{p,t} = 2 \end{cases} \quad (18)$$

with transition matrix

$$P = \begin{bmatrix} \phi_{p,1} & 1 - \phi_{p,2} \\ 1 - \phi_{p,1} & \phi_{p,2} \end{bmatrix} \quad \text{and} \quad P_{ij} = prob(s_{p,t} = i|s_{p,t-1} = j). \quad (19)$$

7.2 Regime Switching in Technological Progress

We can incorporate regime-switching nonlinearity in technological progress by letting average technology growth $γ$ evolve according to a two-state Markov-switching process. Equation
(1) becomes

$$\log A_t = \log \gamma_t + \log A_{t-1} + z_t,$$

(20)

where

$$\gamma_t = \begin{cases} \gamma_L & \text{if } s_{a,t} = 1 \\ \gamma_H & \text{if } s_{a,t} = 2 \end{cases}$$

(21)

with transition matrix

$$\mathcal{P} = \begin{bmatrix} \phi_{a,1} & 1 - \phi_{a,2} \\ 1 - \phi_{a,1} & \phi_{a,2} \end{bmatrix}$$

and

$$P_{ij} = \text{prob}(s_{a,t} = i|s_{a,t-1} = j).$$

(22)

7.3 Empirical Results for Regime Switching

*** In progress.

8 Concluding Remarks
References


### Tables and Figures

#### Table 1: Priors for structural parameters of DSGE model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution</th>
<th>Para (1)</th>
<th>Para (2)</th>
<th>Parameter</th>
<th>Distribution</th>
<th>Para (1)</th>
<th>Para (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau$</td>
<td>Gamma</td>
<td>2</td>
<td>0.5</td>
<td>100$\sigma_r$</td>
<td>InvGamma</td>
<td>0.3</td>
<td>4</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Beta</td>
<td>0.1</td>
<td>0.05</td>
<td>100$\sigma_g$</td>
<td>InvGamma</td>
<td>0.4</td>
<td>4</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Gamma</td>
<td>0.2</td>
<td>0.1</td>
<td>100$\sigma_z$</td>
<td>InvGamma</td>
<td>0.4</td>
<td>4</td>
</tr>
<tr>
<td>$1/g$</td>
<td>Fixed</td>
<td>0.85</td>
<td>N/A</td>
<td>$\rho_{\sigma_r}$</td>
<td>Normal</td>
<td>0.9</td>
<td>0.07</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>Gamma</td>
<td>1.5</td>
<td>0.25</td>
<td>$\rho_{\sigma_g}$</td>
<td>Normal</td>
<td>0.9</td>
<td>0.07</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>Gamma</td>
<td>0.12</td>
<td>0.05</td>
<td>$\rho_{\sigma_z}$</td>
<td>Normal</td>
<td>0.9</td>
<td>0.07</td>
</tr>
<tr>
<td>$\rho_r$</td>
<td>Beta</td>
<td>0.75</td>
<td>0.1</td>
<td>100$\sigma_{\sigma_r}$</td>
<td>InvGamma</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>Beta</td>
<td>0.5</td>
<td>0.2</td>
<td>100$\sigma_{\sigma_g}$</td>
<td>InvGamma</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>Beta</td>
<td>0.5</td>
<td>0.2</td>
<td>100$\sigma_{\sigma_z}$</td>
<td>InvGamma</td>
<td>2.5</td>
<td>4</td>
</tr>
<tr>
<td>400 log($1/\beta$)</td>
<td>Gamma</td>
<td>1</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>400 log $\pi$</td>
<td>Gamma</td>
<td>2.48</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100 log $\gamma$</td>
<td>Normal</td>
<td>0.4</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
1. For the linear DSGE models and the model with stochastic volatility, we fix $\nu$ at 0.1.
2. Para (1) and Para(2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; and $s$ and $\nu$ for the Inverse Gamma distribution, where $p_{IG}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. 

---

20
Table 2: Real-time forecast RMSEs, 1991Q4-2011Q1

<table>
<thead>
<tr>
<th></th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Output Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.682</td>
<td>0.696</td>
<td>0.704</td>
<td>0.723</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.933 (0.017)</td>
<td>0.938 (0.089)</td>
<td>0.948 (0.136)</td>
<td>0.983 (0.396)</td>
</tr>
<tr>
<td>(b) Inflation Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.263</td>
<td>0.265</td>
<td>0.299</td>
<td>0.339</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>1.021 (0.942)</td>
<td>1.032 (0.845)</td>
<td>1.045 (0.821)</td>
<td>1.025 (0.632)</td>
</tr>
<tr>
<td>(c) Fed Funds Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.157</td>
<td>0.262</td>
<td>0.406</td>
<td>0.547</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.944 (0.004)</td>
<td>0.966 (0.148)</td>
<td>1.000 (0.497)</td>
<td>1.047 (0.921)</td>
</tr>
</tbody>
</table>

Notes: 1. RMSEs for benchmark DSGE model in the first panel, RMSE ratios in all others. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. The forecast errors are calculated using actuals that are obtained from the most recent vintage.
3. $p$-values of $t$-tests of equal MSE, taking the linear DSGE models with constant volatilities as the benchmark, are given in parentheses. These are one-sided Diebold-Mariano tests, of the null of equal forecast accuracy against the alternative that the non-benchmark model in question is more accurate. The standard errors entering the test statistics are computed with the Newey-West estimator, with a bandwidth of 0 at the 1-quarter horizon and $n^{1/3}$ in the other cases. $n$ is the number of forecasting origins.
Table 3: Real-time forecast coverage rates, 1991Q4-2011Q1 (70%)

<table>
<thead>
<tr>
<th></th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Output Growth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.949 (0.000)</td>
<td>0.909 (0.000)</td>
<td>0.893 (0.000)</td>
<td>0.914 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.769 (0.147)</td>
<td>0.779 (0.149)</td>
<td>0.800 (0.040)</td>
<td>0.829 (0.106)</td>
</tr>
<tr>
<td><strong>Inflation Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.872 (0.000)</td>
<td>0.857 (0.002)</td>
<td>0.947 (0.000)</td>
<td>0.943 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.782 (0.079)</td>
<td>0.779 (0.135)</td>
<td>0.853 (0.004)</td>
<td>0.857 (0.032)</td>
</tr>
<tr>
<td><strong>Fed Funds Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.936 (0.000)</td>
<td>0.896 (0.000)</td>
<td>0.827 (0.066)</td>
<td>0.800 (0.324)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.846 (0.000)</td>
<td>0.701 (0.986)</td>
<td>0.640 (0.499)</td>
<td>0.600 (0.343)</td>
</tr>
</tbody>
</table>

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. The table reports the frequencies with which actual outcomes fall within 70 percent bands computed from the posterior distribution of forecasts.
3. The table includes in parentheses $p$-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on $t$-statistics using standard errors computed with the Newey-West estimator, with a bandwidth of 0 at the 1-quarter horizon and $n^{1/3}$ in the other cases. $n$ is the number of forecasting origins.
Table 4: **Real-time forecast coverage rates (LR Test), 1991Q4-2011Q1 (70%)**

<table>
<thead>
<tr>
<th></th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Output Growth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.949 (0.000)</td>
<td>0.909 (0.000)</td>
<td>0.893 (0.000)</td>
<td>0.914 (0.002)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.769 (0.171)</td>
<td>0.779 (0.118)</td>
<td>0.800 (0.049)</td>
<td>0.829 (0.080)</td>
</tr>
<tr>
<td>(b) Inflation Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.872 (0.000)</td>
<td>0.857 (0.001)</td>
<td>0.947 (0.000)</td>
<td>0.943 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.782 (0.103)</td>
<td>0.779 (0.118)</td>
<td>0.853 (0.002)</td>
<td>0.857 (0.030)</td>
</tr>
<tr>
<td>(c) Fed Funds Rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.936 (0.000)</td>
<td>0.896 (0.000)</td>
<td>0.827 (0.012)</td>
<td>0.800 (0.180)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.846 (0.003)</td>
<td>0.701 (0.980)</td>
<td>0.640 (0.265)</td>
<td>0.600 (0.209)</td>
</tr>
</tbody>
</table>

*Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008). 2. The table reports the frequencies with which actual outcomes fall within 70 percent bands computed from the posterior distribution of forecasts. 3. The table includes in parentheses $p$-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on the LR test.*
Table 5: **Real-time 70% interval forecast (1-step-ahead) LR test, 1991Q4-2011Q1**

<table>
<thead>
<tr>
<th></th>
<th>Coverage</th>
<th>Independence</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Output Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>30.87 (0.000)</td>
<td>1.92 (0.165)</td>
<td>32.89 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>1.87 (0.171)</td>
<td>0.25 (0.619)</td>
<td>2.65 (0.266)</td>
</tr>
<tr>
<td>(b) Inflation Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>12.85 (0.000)</td>
<td>0.67 (0.413)</td>
<td>13.79 (0.001)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>2.66 (0.103)</td>
<td>0.65 (0.420)</td>
<td>3.80 (0.149)</td>
</tr>
<tr>
<td>(c) Fed Funds Rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>26.97 (0.000)</td>
<td>1.11 (0.292)</td>
<td>28.21 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>9.00 (0.003)</td>
<td>9.98 (0.002)</td>
<td>19.32 (0.000)</td>
</tr>
</tbody>
</table>

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. The table reports the frequencies with which actual outcomes fall within 70 percent bands computed from the posterior distribution of forecasts.
3. The table includes in parentheses p-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on the LR test.
<table>
<thead>
<tr>
<th></th>
<th>Std. Dev.</th>
<th>Mean</th>
<th>AR(1) coef.</th>
<th>LR test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(a) Output Growth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.537 (0.000)</td>
<td>0.194 (0.004)</td>
<td>0.349 (0.026)</td>
<td>53.740 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.827 (0.349)</td>
<td>0.093 (0.493)</td>
<td>0.254 (0.072)</td>
<td>10.610 (0.014)</td>
</tr>
<tr>
<td><strong>(b) Inflation Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.763 (0.324)</td>
<td>0.059 (0.590)</td>
<td>-0.121 (0.492)</td>
<td>10.467 (0.015)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.868 (0.461)</td>
<td>0.130 (0.368)</td>
<td>-0.000 (0.998)</td>
<td>3.915 (0.271)</td>
</tr>
<tr>
<td><strong>(c) Fed Funds Rate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>0.583 (0.000)</td>
<td>-0.074 (0.646)</td>
<td>0.658 (0.000)</td>
<td>76.000 (0.000)</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>0.866 (0.492)</td>
<td>-0.119 (0.593)</td>
<td>0.744 (0.000)</td>
<td>66.514 (0.000)</td>
</tr>
</tbody>
</table>

2. The normalized forecast error is defined as $\Phi^{-1}(z_{t+1})$, where $z_{t+1}$ denotes the PIT of the one-step ahead forecast error and $\Phi^{-1}$ is the inverse of the standard normal distribution function.
3. The first column reports the estimated standard deviation of the normalized error, along with a p-value for a test of the null hypothesis of a standard deviation equal to 1 (computed by a linear regression of the squared error on a constant, using a Newey-West variance with 3 lags.). The second column reports the mean of the normalized error, along with a p-value for a test of the null of a mean of zero (using a Newey-West variance with 5 lags). The third column reports the AR(1) coefficient and its p-value, obtained by estimating an AR(1) model with an intercept (with heteroskedasticity-robust standard errors). The final column reports the $p$-value of Berkowitz’s (2001) likelihood ratio test for the joint null of a zero mean, unity variance, and no [AR(1)] serial correlation.
Table 7: Log Predictive Score, 1991Q4-2011Q1

<table>
<thead>
<tr>
<th></th>
<th>$h = 1Q$</th>
<th>$h = 2Q$</th>
<th>$h = 4Q$</th>
<th>$h = 8Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>-3.99</td>
<td>-4.20</td>
<td>-4.91</td>
<td>-6.55</td>
</tr>
<tr>
<td>Linear+SV</td>
<td>-3.82</td>
<td>-4.66</td>
<td>-5.70</td>
<td>-6.65</td>
</tr>
</tbody>
</table>

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008). 2. Log predictive scores are defined and computed as in Warne et al. (2012). See appendix for detail.
Figure 1: **Estimated Time Varying Standard Deviations**

**Vintage at January 1992**

**Notes:** Posterior means (solid line) and 80% confidence bands (shaded area) of standard deviations of the structural shocks based on the DSGE model with stochastic volatility. Dotted line is the posterior means of the standard deviations of the structural shocks based on the linear DSGE model with constant volatility. Models are estimated at various points in time with the vintage of data indicated.
Figure 2: Real-time forecast RMSEs, 1991Q4-2011Q1

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. The forecast errors are calculated using actuals that are obtained from the most recent vintage.
Figure 3: PITs, 1-Step-Ahead Prediction, 1991Q4-2011Q1

LINEAR DSGE MODEL

LINEAR DSGE MODEL WITH STOCHASTIC VOLATILITY

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. PITs are grouped into five equally sized bins. Under a uniform distribution, each bin should contain 20% of the PITs, indicated by the solid horizontal lines in the figure.
Figure 4: PITs, 4-Step-Ahead Prediction, 1991Q4-2011Q1

**Linear DSGE Model**

**Linear DSGE Model with Stochastic Volatility**

Notes: 1. a) Linear: the linear DSGE model with constant volatility. b) Linear+SV: the DSGE model with stochastic volatility using the method proposed by Justiniano and Primiceri (2008).
2. PITs are grouped into five equally sized bins. Under a uniform distribution, each bin should contain 20% of the PITs, indicated by the solid horizontal lines in the figure.
Figure 5: Log Predictive Score, 1991Q4-2011Q1

1-Step-Ahead

4-Step-Ahead