# Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility

Francis X. Diebold
University of Pennsylvania

Frank Schorfheide
University of Pennsylvania

Minchul Shin
University of Pennsylvania

This Version: May 1, 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 backgroud, we explore whether incorporating stochastic volatility improves DSGE forecasts (point, interval, and density). 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, 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.

# 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. The empirical findings reported in Justiniano and Primiceri (2008), Fernández-Villaverde and Rubio-Ramírez (2013), and Curdia et al. (2014), who also consider fat-tailed shock distributions, indicate that the fit of DSGE models can be improved by allowing for stochastic volatility in the exogenous shock processes. 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 methods for model solution and posterior analysis. In Section 4 we introduce our approach for real-time DSGE forecast analysis with vintage data, describing our dataset and procedure, and providing initial stochastic volatility estimates. In Sections 5, 6 and 7 we evaluate DSGE point, interval and density forecasts, respectively. We conclude in Section 8.

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

[2] See Pichler (2008), Bache et al. (2011), Herbst and Schorfheide (2012), Del Negro and Schorfheide (2013) and Wolters (2015).

## 2 A New Keynesian DSGE Model

Here we present the DSGE model that is used in the subsequent empirical analysis. It is a small-scale New Keynesian model studied by Del Negro and Schorfheide (2013). 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 of government consumption and finances it using lump-sum taxes. In what follows, we are summarizing the log-linearized equilibrium conditions of this economy. Technology $A_t$ evolves according to

$$\log A_t = (\log \gamma)t + \tilde{z}_t. \tag{1}$$

The first part is a deterministic trend component, whereas the second component is an exogenous stochastic process which may be stationary or exhibit a stochastic trend. We define the change in the stochastic component as

$$z_t = \tilde{z}_t - \tilde{z}_{t-1}.$$

To describe the equilibrium conditions, it is convenient to detrend consumption $C_t$ and output $Y_t$ by the level of technology. The detrended variables are defined as $C_t/A_t$ and $Y_t/A_t$, respectively. Even if $\tilde{z}_t$ follows a unit-root process, the model has a steady state in terms of the detrended variables. Henceforth we express all variables in log deviations from steady state values; for example, $c_t = \log(C_t/A_t) - \log c_*$, where $c_*$ is the steady state value of detrended consumption.

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 can be expressed as

$$c_t = \mathbb{E}_t[c_{t+1} + z_{t+1}] - \frac{1}{\tau}(R_t - \mathbb{E}_t[\pi_{t+1}]), \tag{2}$$

where $c_t$ is consumption, $R_t$ is the nominal interest rate, and $\pi_t$ is inflation. The parameter $\tau$ captures the relative degree of risk aversion. The discount factor $\beta$ of the representative household does not appear in the log-linearized Euler equation.

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 only a fraction of the intermediate-goods producers can re-optimize their prices in each period (Calvo mechanism). The final goods producers simply combine the intermediate goods. In equilibrium the inflation in the price of the final good is determined by a New Keynesian Phillips curve:

$$\pi_t = \frac{\iota}{1+\iota\beta}\pi_{t-1} + \frac{\beta}{1+\iota\beta}\mathbb{E}_t[\pi_{t+1}] + \frac{(1-\zeta\beta)(1-\zeta)}{(1+\iota\beta)\zeta}(c_t + \nu_l y_t), \tag{3}$$

where $\zeta$ is the probability with which price setters are able to re-optimize their prices, $\iota$ is the fraction of price setters that index their price to lagged inflation in the event that they are unable to re-optimize, and $\nu_l$ is the inverse labor supply elasticity of the households.

We assume that a fraction of output is used for government consumption. The log-linearized resource constraint takes the form

$$y_t = c_t + g_t, \tag{4}$$

where $g_t$ is an exogenously evolving government spending shock. The central bank sets nominal interest rates in response to inflation deviations and output growth deviations from their respective targets (steady state values):

$$R_t = \rho_R R_{t-1} + (1-\rho_R)\big[\psi_1 \pi_t + \psi_2(y_t - y_{t-1} + z_t)\big] + m_t, \tag{5}$$

where $m_t$ is a monetary policy shock.

We complete the model by specifying the exogenous shock processes,

$$\begin{aligned} m_t &= \epsilon_{R,t}, \quad \epsilon_{R,t} \sim \mathcal{N}(0, \sigma^2_{R,t}), \\ \tilde{z}_t &= \rho_z(1-\varphi)\tilde{z}_{t-1} + \varphi\tilde{z}_{t-2} + \sigma_{z,t}\epsilon_{z,t}, \quad \epsilon_{z,t} \sim \mathcal{N}(0, \sigma^2_{z,t}), \\ g_t &= \rho_g g_{t-1} + \sigma_{g,t}\epsilon_{g,t}, \quad \epsilon_{g,t} \sim \mathcal{N}(0, \sigma^2_{g,t}). \end{aligned} \tag{6}$$

We assume that $\epsilon_{R,t}$, $\epsilon_{z,t}$, and $\epsilon_{g,t}$ are orthogonal at all leads and lags. 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

$$\sigma_{i,t} = \sigma_i e^{\nu_{i,t}}, \quad \nu_{i,t} = \rho_{\sigma_i}\nu_{i,t-1} + \eta_{i,t},$$

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

# 3 Model Solution and Posterior Analysis

Ignoring for a moment the stochastic volatilities of the structural shock innovations $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}, \epsilon_{g,t}]'$, Equations (2)-(6) form a linear rational expectations system that can be solved with a standard algorithm, e.g., Sims (2002). In preliminary work, we also solved the DSGE model with second-order perturbation techniques. However, except in the vicinity of the zero lower bound on nominal interest rate, our New Keynesian model – using a parameterization that fits U.S. data – does not generate any strong nonlinearities. Thus, to simplify the computations, we simply combine the log-linear approximation with the stochastic volatility processes specified above. This leads to a conditionally (given the three volatility processes) linear Gaussian state-space model.

## 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,

$$\begin{aligned} s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \\ \epsilon_t &\sim iid\mathcal{N}(0, Q(\theta)), \end{aligned} \tag{7}$$

where $s_t = [y_t, y_{t-1}, c_t, \pi_t, R_t, mc_t, m_t, g_t, z_t]'$ is a (non-minimal) vector of state variables, $\Phi_1$ is a $n_s \times n_s$ matrix, $\Phi_\epsilon$ is a $n_s \times n_e$ matrix and $Q$ is a $n_e \times n_e$ matrix, where $n_s$ is the number of state variables and $n_e$ is the number of structural shocks. The elements of the coefficient matrices $(\Phi_1(\theta), \Phi_\epsilon(\theta), Q(\theta))$ 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,

$$\begin{aligned} s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \\ \epsilon_t &\sim iid\mathcal{N}(0, Q_t(\theta)). \end{aligned} \tag{8}$$

More specifically, $Q_t(\theta)$ is a diagonal matrix with elements $e^{2\nu_{i,t}}$, where each $\nu_{i,t}$ has its own transition,

$$\begin{aligned} \nu_{i,t} &= \rho_{\sigma_i}\nu_{i,t-1} + \eta_{i,t} \\ \eta_{i,t} &\sim iid\mathcal{N}(0,\ \omega_i^2), \end{aligned} \tag{9}$$

for $i = R, z, g$. Together with a measurement equation, equations (8) and (9) 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

$$\begin{bmatrix} YGR_t \\ INF_t \\ FFR_t \end{bmatrix} = \begin{bmatrix} 100\log\gamma \\ 400\log\pi_* \\ 400\log(\gamma\pi_*/\beta) \end{bmatrix} + \begin{bmatrix} 100(y_t - y_{t-1} + z_t) \\ 400\pi_t \\ 400R_t \end{bmatrix}, \tag{10}$$

where $\pi_*$ denotes the gross target inflation rate (which equals the steady state inflation rate). In slight abuse of notation (changing the definition of $Y$) we write the measurement equation as

$$Y_t = D(\theta) + Z(\theta)s_t. \tag{11}$$

Here $Y_t$ is now the $n \times 1$ vector of observed variables (composed of $YGR_t$, $INF_t$, and $FFR_t$), $D(\theta)$ is an $n \times 1$ vector that contains the DSGE model-implied mean of the observables, $Z(\theta)$

is an $n \times n_s$ matrix that relates the observables to the model states, and $s_t$ is the $n_s \times 1$ state vector.

## 3.3 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) and adapted by Justiniano and Primiceri (2008) to the estimation of linearized DSGE models with stochastic volatility.[3]

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 summarize them in Table 1. For the model with stochastic volatility, we specify prior distributions as:

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

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).

## 3.4 Prediction

We focus on the DSGE model with stochastic volatility. Let $\nu_t = [\nu_{R,t}, \nu_{g,t}, \nu_{z,t}]'$. We generate draws from the posterior predictive density using the decomposition,

$$\begin{aligned} & p(Y_{T+1:T+H}|Y_{1:T}) \\ &= \int_{(\theta, s_T, \nu_T)} \Bigg[ \int_{s_{T+1:T+H}, \nu_{T+1:T+H}} p(Y_{T+1:T+H}|s_{T+1:T+H}) \\ & \quad \times p(s_{T+1:T+H}, \nu_{T+1:T+H}|\theta, s_T, \nu_T, Y_{1:T}) d(s_{T+1:T+H}, \nu_{T+1:T+H}) \Bigg] \\ & \quad \times p(\theta, s_T, \nu_T|Y_{1:T}) d(\theta, s_T, \nu_T). \end{aligned} \tag{12}$$

[3] Detailed descriptions of the posterior simulator can be found in Justiniano and Primiceri (2008) and Del Negro and Schorfheide (2011).

Table 1: PRIORS FOR DSGE MODEL PARAMETERS

| Parameter | Distribution | Para (1) | Para (2) | Parameter | Distribution | Para (1) | Para (2) |
|---|---|---|---|---|---|---|---|
| $\tau$ | Normal | 1.50 | 0.37 | $\rho_R$ | Beta | 0.50 | 0.20 |
| $\nu_l$ | Gamma | 2.00 | 0.75 | $\rho_g$ | Beta | 0.50 | 0.20 |
| $\iota$ | Beta | 0.50 | 0.15 | $\rho_z$ | Beta | 0.50 | 0.20 |
| $\zeta$ | Beta | 0.50 | 0.10 | $\varphi_z$ | Uniform | -1.00 | 1.00 |
| $\psi_1$ | Normal | 1.50 | 0.25 | $100\sigma_R$ | InvGamma | 0.10 | 2.00 |
| $\psi_2$ | Normal | 0.12 | 0.05 | $100\sigma_g$ | InvGamma | 0.10 | 2.00 |
| $400\log(1/\beta)$ | Gamma | 1.00 | 0.40 | $100\sigma_z$ | InvGamma | 0.10 | 2.00 |
| $400\log(\pi_*)$ | Gamma | 2.48 | 0.40 | | | | |
| $100\log(\gamma)$ | Normal | 0.40 | 0.10 | | | | |

Notes: Para (1) and Para(2) contain means and 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 $p_{IG}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. Priors for stochastic volatility are presented in the main text.

We use the subscript $t_1 : t_2$ to indicate sequences from $t_1$ to $t_2$, e.g., $Y_{1:T}$ is shorthand for $Y_1, \dots, Y_T$. The decomposition shows how the predictive density reflects uncertainty about parameters and states at the forecast origin, $p(\theta, s_T, \nu_T|Y_{1:T})$, and uncertainty about future states. Motivated by this decomposition, we generate draws from the predictive density, adapting the algorithm of Del Negro and Schorfheide (2013) to account for the hidden volatility process $\nu_t$.

**Algorithm 1 (Predictive Density Draws)**

*For $j = 1$ to $n_{sim}$,*

1. *Draw $(\theta^{(j)}, s_T^{(j)}, \nu_T^{(j)})$ from the posterior distribution $p(\theta, s_T, \nu_T|Y_{1:T})$.*
2. *Draw from $p(s_{T+1:T+H}, \nu_{T+1:T+H}|\theta^{(j)}, s_T^{(j)}, \nu_T^{(j)})$ as follows:*
   (a) *Draw the sequence of volatility innovations $\eta^{(j)}_{i,T+1:T+H} \sim \mathcal{N}\big(0, (\omega_i^2)^{(j)}\big)$ for $i = R, z, g$.*
   (b) *Starting from $\nu_T^{(j)}$, iterate the volatility law of motion (9) forward to obtain the sequence $\nu^{(j)}_{T+1:T+H}$:*
   $$\nu^{(j)}_{i,t} = \rho^{(j)}_{\sigma_i}\nu^{(j)}_{i,t-1} + \eta^{(j)}_{i,t}, \quad t = T+1, \dots, T+H, \quad i = R, z, g.$$
   (c) *Draw the structural shock innovations $\epsilon^{(j)}_{i,T+1:T+H} \sim \mathcal{N}\big(0, e^{2\nu^{(j)}_{i,t}}\big)$ for $i = R, z, g$.*

(d) *Starting from* $s_T^{(j)}$*, iterate the state transition equation (8) forward:*

$$s_t^{(j)} = \Phi_1(\theta^{(j)}) s_{t-1}^{(j)} + \Phi_\epsilon(\theta^{(j)}) \epsilon_t^{(j)}, \quad t = T+1, ..., T+H.$$

3. *Compute the sequence* $Y_{T+1:T+H}^{(j)}$ *using the measurement equation (11):*

$$Y_t^{(j)} = D(\theta^{(j)}) + Z(\theta^{(j)}) s_t^{(j)}, \quad t = T+1, ..., T+H.$$

Algorithm 1 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 we take 30,000 draws from the posterior distribution $p(\theta, s_T, \nu_T | Y_{1:T})$. We discard the first 10,000 draws and select every 20th draw to get $1,000$ draws of parameters and initial states. For each of these draws, we execute Steps 2 and 3 of the algorithm 20 times, which produces a total of $n_{sim} = 20,000$ draws from the predictive distribution.

# 4 Real-Time DSGE Forecast Analysis with Vintage Data

## 4.1 Empirical Procedure

We evaluate DSGE forecasts using the real-time data set constructed by Del Negro and Schorfheide (2013), who built data vintages aligned with the publication dates of the Blue Chip survey and the Federal Reserve Board's Greenbook, extending the data set compiled by Edge and Gürkaynak (2010). In this paper we use the Del Negro-Schorfheide data set matched to the Blue Chip survey publication dates.

Our first forecast origin is (two weeks prior to) January 1992, and our last forecast origin for one-step-ahead forecasts is (two weeks prior to) April 2011. We use data vintages in April, July, October, and January. Hence we make use of four data vintages per year, for a total of 78. The estimation sample starts from 1964:Q2 for all vintages. We compute forecast errors based on actuals from the most recent vintage, which best estimate the "truth."[4]

To evaluate forecasts we recursively estimate DSGE models over 78 vintages, starting

[4] 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 conclusions regarding DSGE model forecasting performance are generally not affected by the choice of actuals, as did Rubaszek and Skrzypczyński (2008).

from the January 1992 vintage. Hence, for example, for the January 1992 vintage we estimate DSGE models with sample 1964:Q2 - 1991:Q3 and generate forecasts for 1991:Q4 (one step ahead) through 1993:Q2 (eight steps ahead). We then expand the sample gradually, eventually incorporating all vintages from January 1992 through April 2011.

## 4.2 On the Use of 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.[5] 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,

$$p(Y_{t+1}|Y_{1:t}, M_i) = \frac{p(Y_{1:t+1}|M_i)}{p(Y_{1:t}|M_i)},$$

so that

$$\prod_{t=1}^{T-1} p(Y_{t+1}|Y_{1:t}, M_i) = \frac{p(Y_{1:T}|M_i)}{p(Y_1|M_i)}.$$

Hence one can say that Bayesian model selection based on the full-sample predictive performance 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 $Y_{1:t}^{(T)}$ be the data up to time $t$ viewed from the time-$T$ vantage point (vintage $T$), and let $Y_{1:t}^{(t)}$ 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 $\prod_{t=1}^{T-1} p(Y_{t+1}|Y_{1:t}^{(T)}, M_i)$, but rather $\prod_{t=1}^{T-1} p(Y_{t+1}|Y_{1:t}^{(t)}, M_i)$. That is precisely what we implement.

[5] See Diebold (2015).

### 4.3 The Estimated Volatility Path

Our hope, explored subsequently, is that stochastic-volatility DSGE models will produce better forecasts – particularly better interval and density forecasts – than their fixed-volatility counterparts. A necessary condition is that volatility actually be stochastic and indeed highly-variable. Hence we begin by examining and comparing estimated structural shock variances from constant-volatility and stochastic-volatility DSGE models.

In Figure 1 we report posterior-mean stochastic-volatility estimates (solid lines) and constant-volatility estimates (dotted lines) obtained from different real-time data vintages. The vintages are those of January 1992, January 2002, and January 2011, and the corresponding samples end in 1991:Q3, 2001:Q3, and 2010:Q3. The general shapes of volatility are very similar across vintages.

Overall, the estimates confirm significant time variation in volatility. In particular, all volatilities fall sharply with the mid-1980's "Great Moderation." Technology shock volatility, moreover, rises sharply in recent years.

It is interesting to contrast the volatility estimates from the constant-volatility DSGE model. They systematically overstate volatility once the Great Moderation begins, because in significant part the model attempts to fit the high volatility before the Great Moderation.

## 5 Point Forecast Construction and Evaluation

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

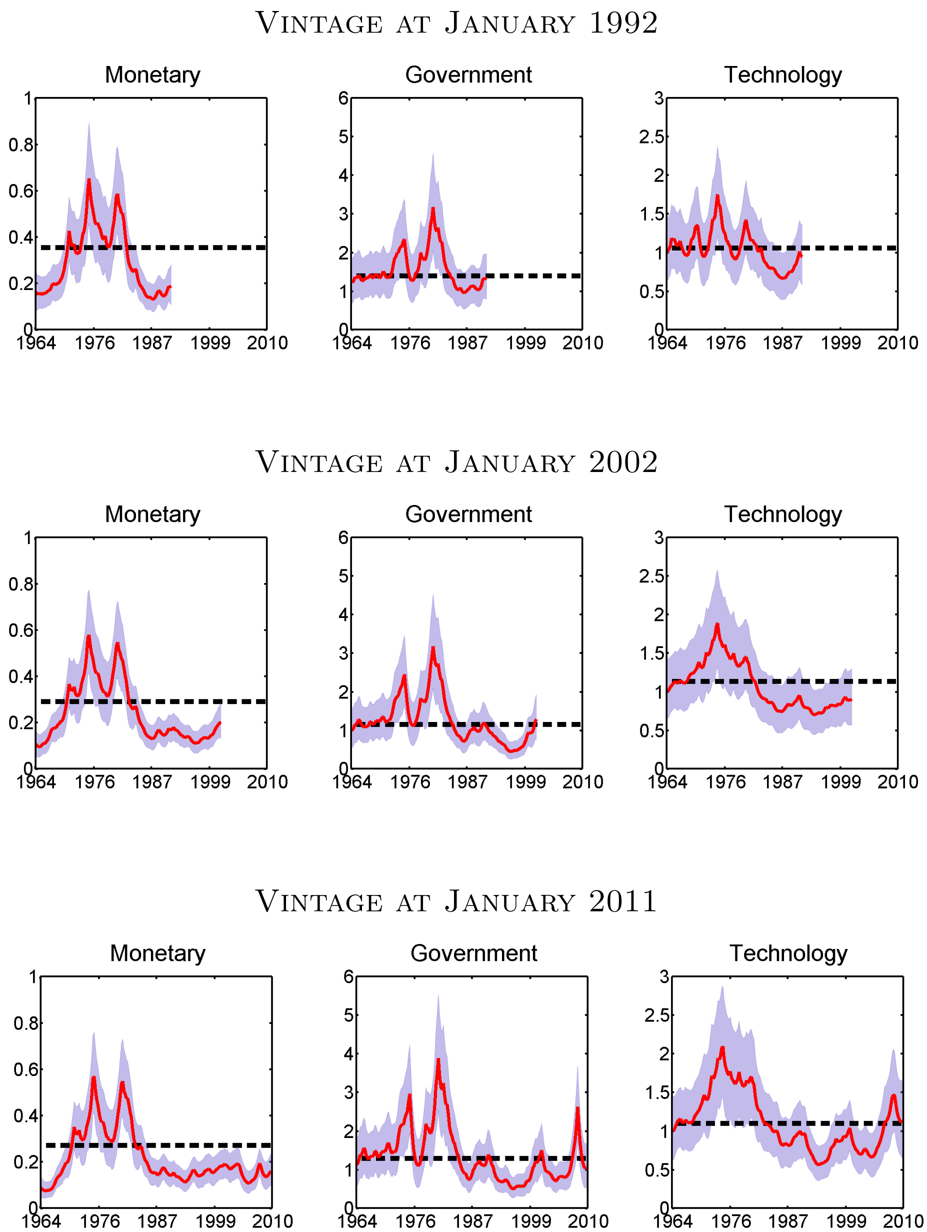
$$\hat{Y}_{T+h|T} = \int_{Y_{T+h}} Y_{T+h}\, p(Y_{T+h}|Y_{1:T}) dY_{T+h} \approx \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} Y_{T+h}^{(j)},$$

where the draws $Y_{T+h}^{(j)}$ are generated with Algorithm 1. The posterior mean is of course the optimal predictor under quadratic loss. To compare the performance of point forecasts we use 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.

Figure 1: Estimated Time-Varying Standard Deviations




Notes: We show estimation results for three different data vintages. We show posterior means (solid line) and 80 percent confidence bands (shaded area) of standard deviations of the structural shocks based on the DSGE model with stochastic volatility. The dotted line is the posterior mean of the standard deviations of the structural shocks based on the linear DSGE model with constant volatility.

Table 2: POINT FORECAST RMSE's

| | $h = 1Q$ | $h = 2Q$ | $h = 4Q$ | $h = 8Q$ |
|---|---|---|---|---|
| | (a) Output Growth | | | |
| Linear | 0.636 | 0.638 | 0.645 | 0.685 |
| Linear+SV | 1.038 | 1.036 | 1.037 | 1.028 |
| | (0.959) | (0.903) | (0.897) | (0.795) |
| | (b) Inflation Rate | | | |
| Linear | 0.250 | 0.273 | 0.310 | 0.354 |
| Linear+SV | 0.971 | 0.974 | 0.969 | 0.969 |
| | (0.031) | (0.037) | (0.076) | (0.095) |
| | (c) Fed Funds Rate | | | |
| Linear | 0.162 | 0.267 | 0.408 | 0.539 |
| Linear+SV | 0.883 | 0.919 | 0.984 | 1.064 |
| | (0.000) | (0.009) | (0.329) | (0.940) |

Notes: The real-time forecast sample is 1991Q4-2011Q1. We calculate forecast errors using actuals obtained from the most recent vintage. We show RMSE's for the benchmark constant-volatility DSGE model in the first line of each panel, and RMSE ratios in the second. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility. In parentheses we show $p$-values of Diebold-Mariano tests of equal MSE against the one-sided alternative that the model with stochastic volatility is more accurate, obtained using standard normal critical values. We compute the standard errors entering the Diebold-Mariano statistics using Newey-West with bandwidth 0 at the 1-quarter horizon and $n^{1/3}$ in the other cases, where $n$ is the number of forecasting origins.

In Table 2 we present real-time forecasts RMSE's for 1991:Q4-2011:Q1.[6] We show RMSE's for the benchmark constant-volatility DSGE model in the first line of each panel, and RMSE ratios in the second.[7] "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility. In parentheses we show $p$-values of Diebold-Mariano tests of equal MSE against the one-sided alternative that the model with stochastic volatility is more accurate.

Stochastic-volatility DSGE forecasts are significantly more accurate for inflation at all horizons, and for fed funds at shorter horizons.[8] In contrast, output growth forecast accu-

[6]We calculate forecast errors using actuals obtained from the most recent vintage.

[7]Ratios less than one indicate that the stochastic-volatility model forecasts are more accurate than benchmark model forecasts.

[8]In general, stochastic-volatility DSGE RMSE's get closer to constant-variance DSGE RMSE's as the forecast horizon increases.

racy is very similar across models and horizons. This basic scenario – incorporating stochastic volatility appears somewhat helpful for point forecasting (presumably due to enhanced parameter estimation efficiency), but not massively helpful – is precisely what one would expect. That is, if stochastic volatility is important, one expects much greater contributions to interval and density forecasting performance, to which we now turn.

# 6 Interval Forecast Construction and Evaluation

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 $\{Y^{(j)}_{i,T+h}\}_{j=1}^{n_{sim}}$.

## 6.1 Relative Evaluation Standards: Coverage and Length

In the interval forecast evaluation that follows, we consider both relative standards (coverage, length) and absolute standards (conditional calibration).

### 6.1.1 Coverage Rates

In Table 3 (A) we report the frequency with which real-time outcomes for output growth, inflation rate, and the federal funds rate fall inside real-time 70-percent highest posterior density intervals.[9] Correct coverage corresponds to frequencies of about 70-percent, whereas a frequency of greater than (less than) 70 percent means that on average over a given sample, the posterior density is too wide (narrow). In parentheses we show $p$-values of $t$-statistics of the hypothesis of correct coverage (empirical = nominal coverage of 70 percent), calculated using Newey-West standard errors.

Table 3 (A) makes clear that the constant-volatility DSGE models tend to be too wide, so that actual outcomes fall inside the intervals much more frequently than the nominal 70-percent rate. For example, for the one-step-ahead forecast horizon, the constant-volatility DSGE model coverage rates range from 87 percent to 94 percent. Based on the reported $p$-values, all empirical departures from 70 percent nominal coverage are statistically significant.

[9] We obtain "actuals" from the most recent vintage. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility.

Table 3: 70 PERCENT INTERVAL FORECAST EVALUATION

| (A) Coverage rate | $h = 1Q$ | $h = 2Q$ | $h = 4Q$ | $h = 8Q$ |
|---|---|---|---|---|
| | | (a) Output Growth | | |
| Linear | 0.923 (0.000) | 0.909 (0.000) | 0.907 (0.000) | 0.914 (0.000) |
| Linear+SV | 0.769 (0.147) | 0.727 (0.630) | 0.747 (0.493) | 0.743 (0.625) |
| | | (b) Inflation Rate | | |
| Linear | 0.872 (0.000) | 0.857 (0.001) | 0.933 (0.000) | 0.914 (0.000) |
| Linear+SV | 0.718 (0.725) | 0.740 (0.522) | 0.747 (0.486) | 0.686 (0.882) |
| | | (c) Fed Funds Rate | | |
| Linear | 0.936 (0.000) | 0.831 (0.021) | 0.787 (0.252) | 0.743 (0.721) |
| Linear+SV | 0.769 (0.147) | 0.584 (0.111) | 0.587 (0.187) | 0.514 (0.060) |

| (B) Interval length | $h = 1Q$ | $h = 2Q$ | $h = 4Q$ | $h = 8Q$ |
|---|---|---|---|---|
| | | (a) Output Growth | | |
| Linear | 2.044 | 2.076 | 2.092 | 2.099 |
| Linear+SV | 1.383 | 1.411 | 1.430 | 1.432 |
| | | (b) Inflation Rate | | |
| Linear | 0.725 | 0.899 | 1.051 | 1.171 |
| Linear+SV | 0.516 | 0.623 | 0.712 | 0.777 |
| | | (c) Fed Funds Rate | | |
| Linear | 0.531 | 0.706 | 0.967 | 1.269 |
| Linear+SV | 0.284 | 0.401 | 0.580 | 0.786 |

Notes: The real-time forecast sample is 1991Q4-2011Q1. We obtain "actuals" from the most recent vintage. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility. In Panel (A) we report the frequencies with which outcomes fall in 70-percent bands computed from the posterior predictive density. In parentheses we show $p$-values of $t$-statistics of the hypothesis of correct coverage (empirical = nominal coverage of 70 percent), calculated using Newey-West standard errors with bandwidth 0 at the 1-quarter horizon and $n^{1/3}$ in the other cases, where $n$ is the number of forecasting origins. In Panel (B) we report the average lengths of prediction intervals.

The coverage of the stochastic-volatility DSGE intervals, in contrast, is strikingly good. For all variables and horizons, estimated coverage is much closer to 70 percent, and the $p$-values indicate that in the vast majority of cases any deviation is statistically insignificant.

### 6.1.2 Interval Length

In Table 3 (B) we show average prediction interval lengths. Average lengths based on the stochastic-volatility model are roughly 30 percent shorter than those from the constant-volatility model. Hence the stochastic-volatility intervals dominate on *both* the coverage and length dimensions.

## 6.2 Absolute Evaluation Standards: Conditional Calibration

We also consider an absolute standard for interval forecasts, 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.

In Table 4 we present results of Christoffersen tests for 70-percent 1-step-ahead interval forecasts, 1991Q4-2011Q1. We show separate and joint tests for correct coverage and independence. The coverage tests consistently find no flaws in the stochastic-volatility DSGE intervals, while simultaneously consistently finding severe flaws in the constant-volatility DSGE intervals.

Neither the stochastic-volatility nor the constant-volatility DSGE interval forecasts perform consistently well in terms of the independence test. This is precisely as expected, however, because small-scale DSGE models are well-known to have weak propagation mechanisms that fail to fully capture the conditional-mean dependence (serial correlation) in macroeconomic time series.[10] Incorporating stochastic volatility can naturally fix mis-calibration problems, but there is no way for it to fix inadequate conditional-mean dynamics.

Finally, neither the stochastic-volatility nor the constant-volatility DSGE interval forecasts perform consistently well in terms of the joint test. Presumably the stochastic-volatility intervals fail the joint test due to their failing independence, whereas the constant-volatility intervals fail the joint test due to their failing *both* correct coverage and independence.

[10]See, for example, Cogley and Nason (1995).

Table 4: Christoffersen Tests

| | Coverage | Independence | Joint |
|---|---|---|---|
| (a) Output Growth | | | |
| Linear | 23.50 (0.000) | 16.28 (0.000) | 39.95 (0.000) |
| Linear+SV | 1.87 (0.171) | 1.23 (0.267) | 3.63 (0.163) |
| (b) Inflation Rate | | | |
| Linear | 12.85 (0.000) | 0.45 (0.503) | 13.57 (0.001) |
| Linear+SV | 0.12 (0.728) | 2.76 (0.096) | 3.55 (0.169) |
| (c) Fed Funds Rate | | | |
| Linear | 26.97 (0.000) | 1.11 (0.292) | 28.21 (0.000) |
| Linear+SV | 1.87 (0.171) | 5.31 (0.021) | 7.71 (0.021) |

Notes: We show results for 70-percent 1-step-ahead interval forecasts. The real-time forecast sample is 1991Q4-2011Q1. We obtain "actuals" from the most recent vintage. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility. We show tests for coverage and for independence, as well as joint tests, with $p$-values in parentheses.

# 7 Density Forecast Construction and Evaluation

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

## 7.1 Relative Evaluation Standards: Log Predictive Likelihood

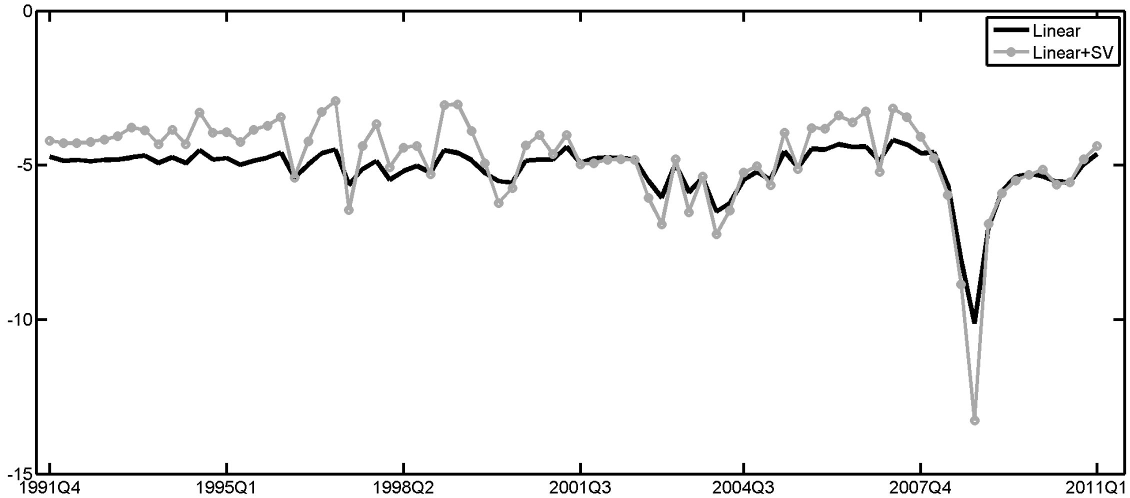
We use the log predictive likelihood for relative density forecast accuracy comparison, as in Warne et al. (2012).[11] The predictive likelihood is

$$S_M(h) = \frac{1}{P-h} \sum_{T=R}^{R+P-h} \log p(Y_{T+h}|Y_{1:T}), \quad h = 1, 2, ..., H, \tag{13}$$

where $R$ is the starting point of the forecast evaluation sample, $P$ is the number of forecast origins, and $h$ is the forecast horizon. $M$ denotes marginal, as opposed to joint, predictive

[11]We will often refer simply to the "predictive likelihood," with the understanding that logs have been taken.

Figure 2: 1-Step-Ahead Predictive Densities




Notes: The real-time forecast sample is 1991Q4-2011Q1.

likelihood, which can be defined as

$$S_J(h) = \frac{1}{P-h} \sum_{T=R}^{R+P-h} \log p(Y_{T+1:T+h}|Y_{1:T}), \quad h = 1, 2, ..., H.$$

Obviously the joint and marginal predictive likelihood concepts lead to the same quantity when $h = 1$. To compute the marginal predictive density $S_M(h)$, after Step 3 of Algorithm 1 we evaluate the density $p(Y_{T+h}|\nu^{(j)}_{T+1:T+h}, \theta^{(j)}, s^{(j)}_T, \nu^{(j)}_T)$. This density is Gaussian and can be obtained from the Kalman filter, treating the observations $Y_{T+1:T+h-1}$ as missing. Averaging across draws $j$ leads to the Monte Carlo approximation

$$p(Y_{T+h}|Y_{1:T}) \approx \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} p(Y_{T+h}|\nu^{(j)}_{T+1:T+h}, \theta^{(j)}, s^{(j)}_T, \nu^{(j)}_T). \tag{14}$$

In Figure 2 we show a time-series plot of 1-step-ahead predictive density values averaged across GDP growth, inflation rate and interest rate forecasts. Those for the stochastic-volatility DSGE model clearly tend to be higher than those for the constant-volatility model.

In Table 5 we present marginal predictive likelihoods for density forecasts at horizons $h = 1, 2, 4, 8$. For most variables and horizons, the stochastic-volatility DSGE model has higher predictive likelihood. The only exceptions are at long horizons (GDP growth at

Table 5: MARGINAL PREDICTIVE LIKELIHOODS $S_M(h)$

| | $h = 1Q$ | $h = 2Q$ | $h = 4Q$ | $h = 8Q$ |
|---|---|---|---|---|
| | (a) Output Growth | | | |
| Linear | -1.14 | -1.15 | **-1.15** | **-1.18** |
| Linear+SV | **-1.03** | **-1.08** | -1.17 | -1.32 |
| | (b) Inflation Rate | | | |
| Linear | -1.93 | -1.95 | -1.98 | -2.04 |
| Linear+SV | **-1.72** | **-1.74** | **-1.80** | **-1.94** |
| | (c) Fed Funds Rate | | | |
| Linear | -2.19 | -2.26 | -2.33 | **-2.40** |
| Linear+SV | **-2.01** | **-2.11** | **-2.26** | -2.47 |

Notes: The real-time forecast sample is 1991Q4-2011Q1. We calculate forecast errors using actuals obtained from the most recent vintage. We present predictive likelihoods for density forecasts at horizons $h = 1, 2, 4, 8$, for output growth, the inflation rate, and the fed funds rate. We show in bold the "winners," for each horizon and each variable.

horizons $h = 4, 8$, and the fed funds rate at horizon $h = 8$).

## 7.2 Absolute Evaluation Standards: 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 $Y_{i,T+h}$,

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

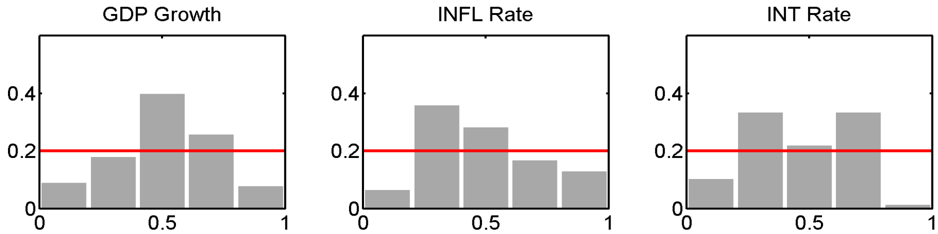
We compute PIT's by the Monte Carlo average of the indicator function,

$$z_{i,h,T} \approx \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} \mathcal{I}\{Y_{i,T+h}^{(j)} \leq Y_{i,T+h}\}.$$

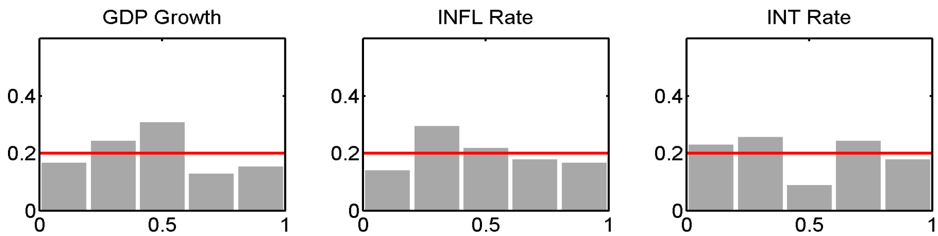
Figure 3: PIT Histograms

**1-Step-Ahead Prediction**

Linear DSGE Model

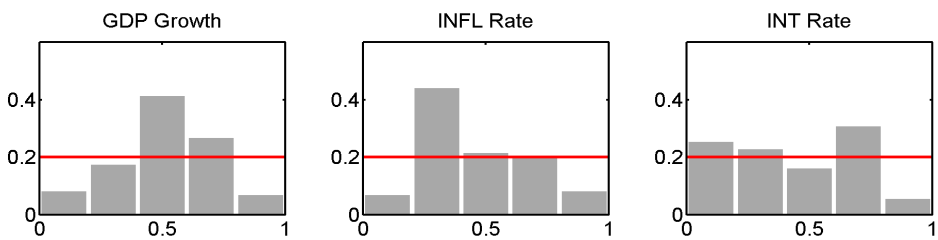



Linear DSGE Model with Stochastic Volatility

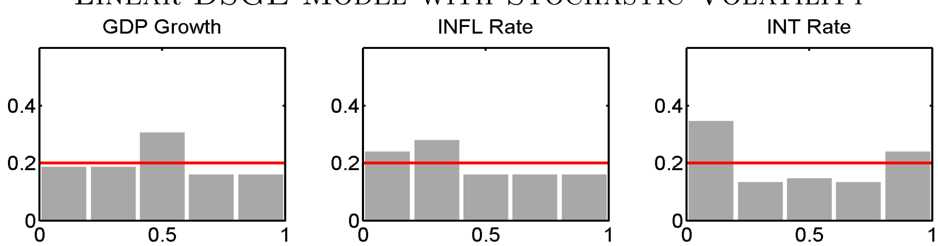



**4-Step-Ahead Prediction**

Linear DSGE Model




Linear DSGE Model with Stochastic Volatility

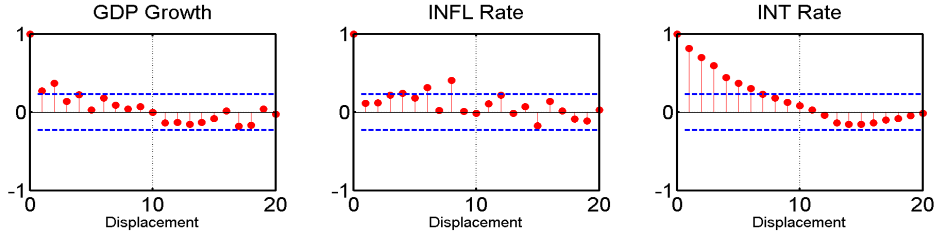



Notes: The real-time forecast sample is 1991Q4-2011Q1. We calculate forecast errors using actuals obtained from the most recent vintage. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility. We group PIT's into five equally-sized bins. Under uniformity, each bin should contain 20 percent of the PIT's, as indicated by the horizontal red lines.

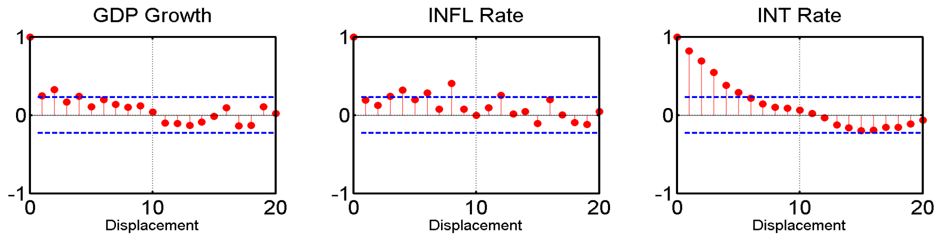
Figure 4: PIT Autocorrelations

**1-Step-Ahead Prediction**

Linear DSGE Model

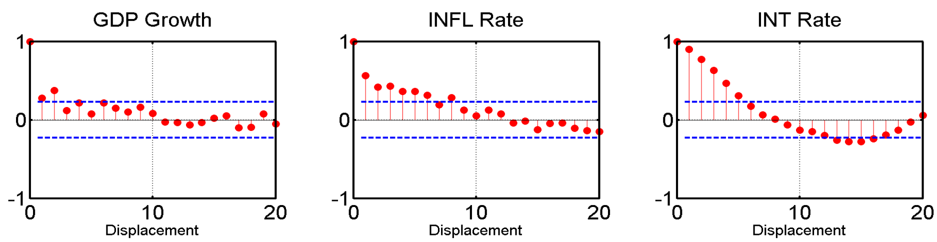



Linear DSGE Model with Stochastic Volatility

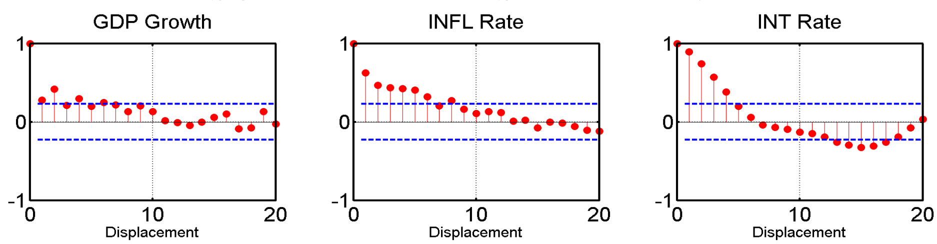



**4-Step-Ahead Prediction**

Linear DSGE Model




Linear DSGE Model with Stochastic Volatility




Notes: The real-time forecast sample is 1991Q4-2011Q1. We calculate forecast errors using actuals obtained from the most recent vintage. "Linear" is the model with constant volatility, and "Linear+SV" is the model with stochastic volatility.

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.

In Figure 3 we report PIT histograms for forecast horizons $h = 1, 4$, for DSGE models with and without stochastic volatility. We group PIT's into five equally sized bins. Under uniformity, each bin should contain 20 percent of the PIT's, as indicated by the horizontal red lines in the figure. Checking histograms alone essentially effectively amounts to checking unconditional calibration.

Histograms for the constant-volatility model appear highly non-uniform. For output growth, too few PIT's are in the extreme bins, indicating that the predictive distribution tends to be too diffuse. Similarly, for the inflation rate, too few PIT's are in the extreme left-tail bin (0-0.2), and for the fed funds rate too few PIT's are in the extreme right-tail bin (0.8-1). In contrast, histograms for the stochastic-volatility model appear much more uniform.

We present PIT sample autocorrelations in Figure 4. Clear deviations from independence are apparent for both the constant-volatility and stochastic-volatility models. Hence, although the stochastic-volatility DSGE models appear correctly unconditionally calibrated (in contrast to the stochastic-volatility models), they are nevertheless not correctly *conditionally* calibrated, because they fail the independence condition. This pattern, and its underlying reasons, matches precisely our earlier results for interval forecasts.

# 8 Conclusion

We have examined the real-time accuracy of point, interval and density forecasts of output growth, inflation, and the federal funds rate, generated from DSGE models with and without stochastic volatility. The stochastic-volatility versions are superior to the constant-volatility versions. We traced the superiority of stochastic-volatility forecasts to superior coverage rates (for interval forecasts) and superior PIT uniformity (for density forecasts) – essentially superior *unconditional* calibration of the stochastic-volatility forecasts. Neither model, however, appears correctly *conditionally* calibrated, as correct conditional calibration requires both correct unconditional calibration and a type of “error independence” condition, which fails to hold.

# References


Bache, W. I., S.A. Jore, J. Mitchell, and S.P. Vahey (2011), “Combining VAR and DSGE Forecast Densities,” *Journal of Economic Dynamics and Control*.

Bloom, N. (2009), “The Impact of Uncertainty Shocks,” *Econometrica*, 77, 623–685.

Christoffersen, P.F. (1998), “Evaluating Interval Forecasts,” *International Economic Review*, 39, 841–862.

Clark, T.E. (2011), “Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility,” *Journal of Business and Economic Statistics*, 29, 327–341.

Cogley, T. and J.M. Nason (1995), “Output Dynamics in Real Business Cycle Models,” *American Economic Review*, 3, 492–511.

Curdia, Vasco, Marco Del Negro, and Daniel Greenwald (2014), “Rare Large Shocks and Great Recessions,” Tech. rep.

Del Negro, M. and F. Schorfheide (2011), “Bayesian Macroeconometrics,” In J. Geweke, G. Koop, and H. van Dijk (eds.) *The Oxford Handbook of Bayesian Econometrics*, Oxford University Press, 293-389.

Del Negro, M. and F. Schorfheide (2013), “DSGE Model-Based Forecasting,” in “Handbook of Economic Forecasting,” (edited by Elliott, G. and A. Timmermann), 2a, 57–140, Elsevier.

Diebold, F.X. (2015), “Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests (with discussion),” *Journal of Business and Economic Statistics*, 33, 1–24.

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

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

Fernández-Villaverde, J. and J.F. Rubio-Ramírez (2007), “Estimating Macroeconomic Models: A Likelihood Approach,” *Review of Economic Studies*, 74, 1059–1087.

Fernández-Villaverde, J. and J.F. Rubio-Ramírez (2013), "Macroeconomics and Volatility: Data, Models, and Estimation," in "Advances in Economics and Econometrics: Tenth World Congress," (edited by Acemoglu, D., M. Arellano, and E. Dekel), 3, 137–183, Cambridge University Press.

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

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

Kim, S., N. Shephard, and S. Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison With ARCH Models," *The Review of Economic Studies*, 65, 361–393.

Pichler, P. (2008), "Forecasting with DSGE Models: The Role of Nonlinearities," *The B.E. Journal of Macroeconomics*, 8, 20.

Primiceri, G.E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–852.

Rubaszek, M. and P. Skrzypczyński (2008), "On the Forecasting Performance of a Small-Scale DSGE Model," *International Journal of Forecasting*, 24, 498–512.

Sims, C.A. and T. Zha (2006), "Were There Regime Switches in U.S. Monetary Policy?" *American Economic Review*, 96, 54–81.

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

Warne, A., G. Coenen, and K. Christoffel (2012), "Forecasting with DSGE-VAR Models," Manuscript, Directorate General Research, European Central Bank.

Wolters, M.H. (2015), "Evaluating Point and Density Forecasts of DSGE Models," *Journal of Applied Econometrics*, 30, 74–96.