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 Illinois

This Version: August 18, 2016

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

Correspondence: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: `fdiebold@sas.upenn.edu` (Francis X. Diebold). Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: `schorf@ssc.upenn.edu` (Frank Schorfheide). Department of Economics, University of Illinois, 214 David Kinley Hall, 1407 W. Gregory Dr., Urbana, IL 61801. Email: `mincshin@illinois.edu` (Minchul Shin).

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. Our empirical findings are very

[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).

similar to those of Clark (2011): the inclusion of stochastic volatility improves predictions in terms of coverage probabilities of interval forecasts, predictive likelihood values, and coverage probabilities of density forecasts.

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. An Online Appendix contains two robustness exercises that involve the evaluation of 90% probability interval forecasts and the evaluation of forecasts based on a pre-Great Recession sample that ends in 2007:Q4.

2 A New Keynesian DSGE Model

Here we present the DSGE model that is used in the subsequent empirical analysis. It is similar to the 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 π_t is inflation. The parameter τ captures the relative degree of risk aversion. The discount factor β 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 households, 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 ζ is the probability with which price setters are able to re-optimize their prices, ι is the fraction of price setters that index their price to lagged inflation in the event that they are unable to re-optimize, and ν_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 and output growth deviations from their respective targets:

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

where m_t is a monetary policy shock and $\pi_{*,t}$ is a central bank's inflation target rate in log-deviation from its long-run mean $\log(\pi_*)$.

The target inflation rate evolves as a stationary AR(1) process with a homoscedastic innovations:

$$\pi^*_t = \rho_{\pi^*}\pi^*_{t-1} + \epsilon_{\pi^*,t}, \quad \epsilon_{\pi^*,t} \sim \mathcal{N}(0, \sigma^2_{\pi^*}). \tag{6}$$

The parameter ρ_{π^*} is expected to be close to one so that π^*_t captures low frequency changes in inflation that we attribute to slowly evolving changes in monetary policy regimes. The time-varying target rate is empirically supported by US data. It mainly captures the fact

that monetary policy was characterized by a shift to a high-inflation period in the 1970s which ended with Volcker's stabilization policy. In the forecasting context, the time-varying target rate captures low frequency shifts in the level of inflation.

We complete the model by specifying laws of motion for the remaining exogenous shock processes:

$$\begin{aligned} m_t &= \epsilon_{R,t}, \quad \epsilon_{R,t} \sim \mathcal{N}(0, \sigma_{R,t}^2), \\ \tilde{z}_t &= \rho_z(1-\varphi)\tilde{z}_{t-1} + \varphi\tilde{z}_{t-2} + \epsilon_{z,t}, \quad \epsilon_{z,t} \sim \mathcal{N}(0, \sigma_{z,t}^2), \\ g_t &= \rho_g g_{t-1} + \epsilon_{g,t}, \quad \epsilon_{g,t} \sim \mathcal{N}(0, \sigma_{g,t}^2). \end{aligned} \tag{7}$$

We assume that $\epsilon_{R,t}$, $\epsilon_{z,t}$, $\epsilon_{g,t}$, and $\epsilon_{\pi^*,t}$ 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

$$\sigma_{i,t} = \sigma_i e^{\nu_{i,t}}, \quad \nu_{i,t} = \rho_{\sigma_i}\nu_{i,t-1} + \eta_{i,t}, \quad i \in \{R, z, g\} \tag{8}$$

where $\eta_{i,t}$ and $\epsilon_{j,t}$ are independent of each other at all leads and lags for all i and j.

3 Model Solution and Posterior Analysis

Ignoring for a moment the stochastic volatilities of the structural shock innovations $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}, \epsilon_{g,t}, \epsilon_{\pi^*,t}]'$, Equations (2)-(7) 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 the 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

A first-order perturbation solution results in a linear transition equation for the state variables,

$$\begin{aligned} s_t &= \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \\ \epsilon_t &\sim iid\mathcal{N}(0, Q(\theta)), \end{aligned} \tag{9}$$

where $s_t = [y_t, y_{t-1}, c_t, \pi_t, R_t, mc_t, m_t, g_t, z_t, \pi_t^*]'$ is a (non-minimal) vector of state variables, Φ_1 is a $n_s \times n_s$ matrix, Φ_ϵ 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 θ.

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,

$$\begin{aligned} s_t &= \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \\ \epsilon_t &\sim iid\mathcal{N}(0, Q_t(\theta)). \end{aligned} \tag{10}$$

More specifically, $Q_t(\theta)$ is a diagonal matrix. The first three diagonal elements are $\sigma_i^2 e^{2\nu_{i,t}}$ for $i \in \{R, z, g\}$. The fourth diagonal element is $\sigma_{\pi^*}^2$. The $\nu_{i,t}$'s have their own transition,

$$\begin{aligned} \nu_{i,t} &= \rho_{\sigma_i}\nu_{i,t-1} + \eta_{i,t} \\ \eta_{i,t} &\sim iid\mathcal{N}(0,\ \sigma_{\sigma_i}^2). \end{aligned} \tag{11}$$

Together with a measurement equation, (10) and (11) 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 GDP growth rates (YGR) and inflation rates (INF), quarterly nominal interest (federal funds) rates (FFR), and 10-year inflation expectation (INF^{10y}) from the Survey of Professional Forecasters maintained by FRB Philadelphia[3]. We measure INF, FFR, and INF^{10y} as annualized percentages, and we measure YGR as a quarterly percentage. We assume that there is no measurement error.

[3]To obtain longer inflation expectation series, we take inflation expectations from the Livingston Survey and the Blue Chip Economic Indicators for the period 1979–1991 and from the Survey of Professional Forecasters (SPF) afterwards. Inflation expectations in this survey are for the CPI, while inflation rates in our estimation and prediction are for the GDP deflator. To correct for this difference, we subtract the average difference between CPI and GDP inflation from the beginning of the sample to the initial point for our forecasting exercise. See Del Negro and Schorfheide (2013) for details.

Then the measurement equation is

$$\begin{bmatrix} YGR_t \\ INF_t \\ FFR_t \\ INF_t^{10y} \end{bmatrix} = \begin{bmatrix} 100 \log \gamma \\ 400 \log \pi_* \\ 400 \log(\gamma \pi_* / \beta) \\ 400 \log \pi_* \end{bmatrix} + \begin{bmatrix} 100\big(y_t - y_{t-1} + z_t\big) \\ 400\pi_t \\ 400R_t \\ 400\mathbb{E}_t \left[\frac{1}{40} \sum_{k=1}^{40} \pi_{t+k}\right] \end{bmatrix}. \tag{12}$$

We link the observed 10-year inflation expectation to the model-implied 10-year inflation expectation in the last line. Our 10-year inflation expectations data start in 1979. Prior to this date, we treat the expectations data as missing and adjust the measurement equation accordingly. As discussed in Del Negro and Schorfheide (2013), the expectations data help to identify the time-varying target rate π_t^* in real time and ensure that the post-1992 inflation forecasts are not contaminated by reversion to a mean that reflects the high inflation rates in the 1970s.

In slight abuse of notation (changing the definition of Y) we write the measurement equation as

$$Y_t = D_t(\theta) + Z_t(\theta) s_t. \tag{13}$$

Here Y_t is now the $n_t \times 1$ vector of observed variables (composed of YGR_t, INF_t, FFR_t, and INF_t^{10y}), $D_t(\theta)$ is an $n_t \times 1$ vector that contains the DSGE model-implied mean of the observables, $Z_t(\theta)$ is an $n_t \times n_s$ matrix that relates the observables to the model states, and s_t is the $n_s \times 1$ state vector. The dimension of the measurement equation deterministically changes over time depending on the availability of the 10-year expectation data.

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.[4]

[4]Detailed descriptions of the posterior simulator can be found in Justiniano and Primiceri (2008), Del Negro and Schorfheide (2011), and Del Negro and Primiceri (2015).

Table 1: Priors for DSGE model parameters

Parameter	Distribution	Para (1)	Para (2)	Parameter	Distribution	Para (1)	Para (2)
τ	Normal	1.50	0.37	ρ_R	Beta	0.50	0.20
ν_l	Gamma	2.00	0.75	ρ_g	Beta	0.50	0.20
ι	Beta	0.50	0.15	φ_z	Uniform	-1.00	1.00
ζ	Beta	0.50	0.10	$100\sigma_R$	InvGamma	0.10	2.00
ψ_1	Normal	1.50	0.25	$100\sigma_g$	InvGamma	0.10	2.00
ψ_2	Normal	0.12	0.05	$100\sigma_z$	InvGamma	0.10	2.00
$400\log(1/\beta)$	Gamma	1.00	0.40	ρ_{π^*}	Beta	0.50	0.20
$400\log(\pi_*)$	Gamma	2.48	0.40	$100\sigma_{\pi^*}$	InvGamma	0.10	2.00
$100\log(\gamma)$	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 ν 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 fix $\rho_z = 1$.

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 and summarize them in Table 1. We fix $\rho_z = 1$, imposing a unit root in technology. For the model with stochastic volatility, we consider two specifications. The first specification follows Justiniano and Primiceri (2008) and assumes that log volatility evolves as random walk

$$\text{SV-RW}: \quad \nu_{i,t} = \nu_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}(0, \sigma^2_{\sigma_i}), \tag{14}$$

where we set the autoregressive parameter to one, $\rho_{\sigma_i} = 1$. For this specification, we impose the inverse gamma prior on $\sigma^2_{\sigma_i}$:

$$\sigma^2_{\sigma_i} \sim \mathcal{IG}(2, 0.0001). \tag{15}$$

This prior specification implies that with 90% probability the standard deviation of a structural shocks can be 18% smaller or 22% larger at the end of the sample in 2011:Q1 compared to its initial level in 1964:Q2.

The second volatility specification relaxes the random walk assumption and assumes the following AR(1) log volatility process:

$$\text{SV-AR}: \quad \nu_{i,t} = c_{\sigma_i} + \rho_{\sigma_i}\nu_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim \mathcal{N}(0, \sigma^2_{\sigma_i}). \tag{16}$$

Here we reparameterized the volatility process in terms of $c_{\sigma_i} = (1 - \rho_{\sigma_i}) \log \sigma_i$. For this specification, we use the following prior distributions:

$$c_{\sigma_i} \sim \mathcal{N}(0, 10), \quad \rho_{\sigma_i} \sim \mathcal{N}(0.9, 0.07), \quad \sigma^2_{\sigma_i} \sim \mathcal{IG}(2, 0.05).$$

We constrain the priors for the $AR(1)$ stochastic-volatility coefficients to be in the stationary

region, $\rho_{\sigma_i} \in (-1, 1)$. This specification is less restrictive than the first one in terms of the *a priori* likely volatility range. The prior implies that standard deviations of structural shocks can be 50% smaller or 100% larger with 90% chance at any given point in time.

Finally, we also consider a specification with a deterministic break in the standard deviation of the structural shocks. We simply assume that structural break happened during the Great Moderation at the end of 1984:

$$\text{DV-SB}: \quad \sigma_{i,t} = \begin{cases} \sigma_{i,0} & \text{if } t \leq 1984:Q4 \\ \sigma_{i,1} & \text{if } t > 1984:Q4, \end{cases} \tag{17}$$

where we estimate $\sigma_{i,0}$ and $\sigma_{i,1}$ separately. We impose the same inverse Gamma prior distribution for $\sigma_{i,0}$ and $\sigma_{i,1}$ with the same parameter values as in the model with constant volatility.

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,

$$\begin{aligned} & p(Y_{T+1:T+H}|Y_{1:T}) \\ &= \int_{(\theta, s_T, \nu_T)} \Bigg[\int_{s_{T+1:T+H}, \nu_{T+1:T+H}} p(Y_{T+1:T+H}|s_{T+1:T+H}) \\ &\quad \times p(s_{T+1:T+H}, \nu_{T+1:T+H}|\theta, s_T, \nu_T, Y_{1:T}) d(s_{T+1:T+H}, \nu_{T+1:T+H}) \Bigg] \\ &\quad \times p(\theta, s_T, \nu_T|Y_{1:T}) d(\theta, s_T, \nu_T). \end{aligned} \tag{18}$$

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 $Y_1, \ldots, Y_T$. 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 ν_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 $\eta_{i,T+1:T+H}^{(j)} \sim \mathcal{N}\big(0, (\omega_i^2)^{(j)}\big)$ for $i = R, z, g$.*

   (b) *Starting from $\nu_T^{(j)}$, iterate the volatility law of motion (11) forward to obtain the sequence $\nu_{T+1:T+H}^{(j)}$:*

$$\nu_{i,t}^{(j)} = \rho_{\sigma_i}^{(j)} \nu_{i,t-1}^{(j)} + \eta_{i,t}^{(j)}, \quad t = T+1, \ldots, T+H, \quad i = R, z, g.$$

   (c) *Draw the structural shock innovations $\epsilon_{i,T+1:T+H}^{(j)} \sim \mathcal{N}\big(0, \sigma_i^{2(j)} e^{2\nu_{i,t}^{(j)}}\big)$ for $i = R, z, g$ and $\epsilon_{\pi^*,T+1:T+H}^{(j)} \sim \mathcal{N}\big(0, \sigma_{\pi^*}^{2(j)}\big)$.*

   (d) *Starting from $s_T^{(j)}$, iterate the state transition equation (10) 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 (13):*

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

Algorithm 1 produces n_{sim} trajectories $Y^{(j)}_{T+1:T+H}$ 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 10th 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 10 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. The survey is conducted over two days, typically beginning on the first or second business day of each month, and then published a few days later. We consider the April, July, October, and January publication dates. The timing of the survey implies that, say, the April 2005 forecast utilizes the "third" estimate of GDP for the fourth quarter of 2004, but not the advance estimate for the first quarter of 2005.

Our first forecast origin is January 1992, and our last forecast origin for one-step-ahead forecasts is April 2011. We recursively estimate the DSGE models for the resulting 78 vintages. The estimation sample starts in 1964:Q2 for all vintages. For example, for the January 1992 vintage we estimate DSGE models based on the sample 1964:Q2 - 1991:Q3

and generate forecasts for 1991:Q4 (one step ahead) through 1993:Q2 (eight steps ahead).[5] We then expand the sample gradually, eventually incorporating all vintages from January 1992 through April 2011, implying that the last observation in the estimation sample is dated 2010:Q4. We compute forecast errors based on actuals from the most recent vintage, which best estimates the "truth."[6] The last observation that we forecast is 2011:Q1. As a robustness check we present forecast evaluation statistics based on samples ending in 2007:Q4, excluding data from the Great Recession. The key findings are very similar to those presented subsequently for the longer sample.

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. 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.[7]

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

[5] At the end of December 1991 the 1991:Q4 NIPA data were not yet available.

[6] 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).

[7] See Diebold (2015).

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.

4.3 The Estimated Volatility Paths

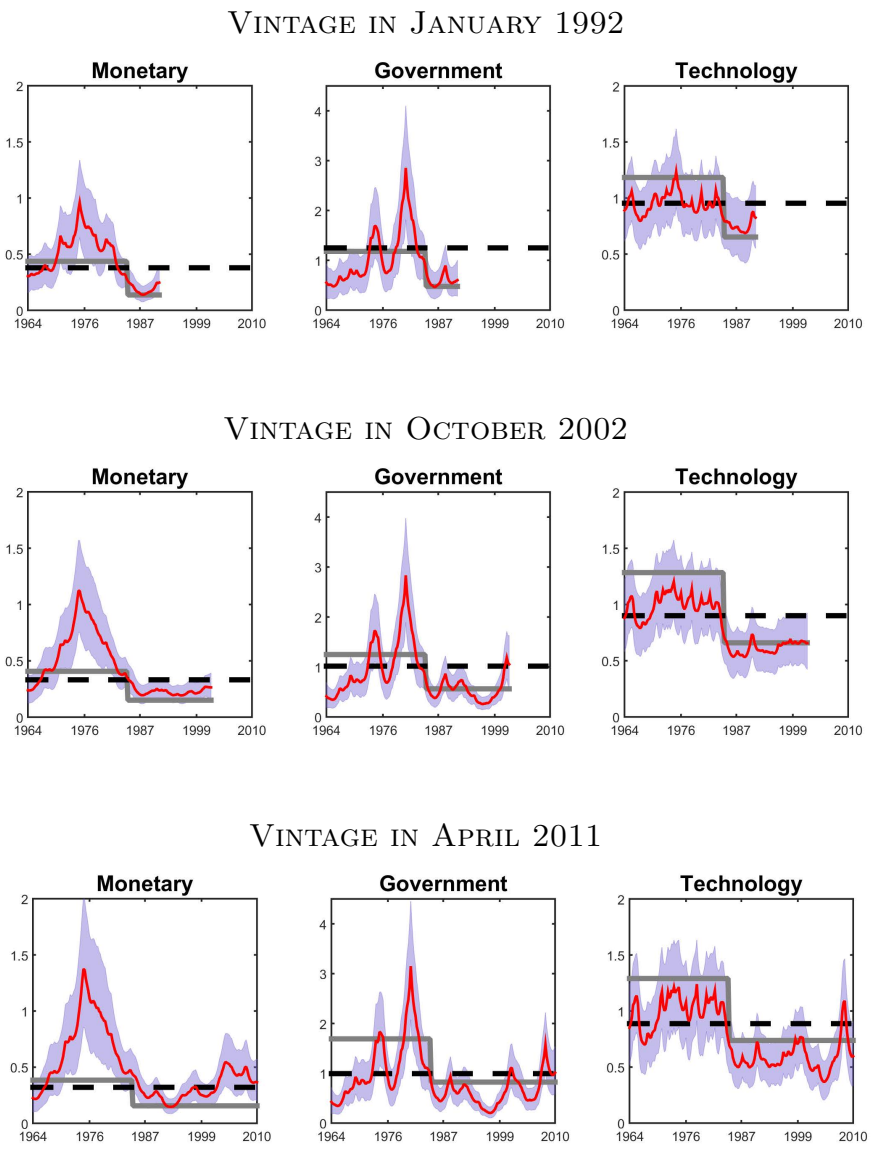
Our hope, explored subsequently, is that stochastic-volatility DSGE models will produce better forecasts – particularly better interval and density forecasts – than their fixed-volatility counterparts. A necessary condition is that volatility actually be stochastic and indeed highly-variable. Hence we 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 for the SV-AR specification (solid lines), constant-volatility estimates (dashed black lines), and estimates based on the structural-break specification DV-SV (solid grey lines) obtained from three different real-time data vintages. The vintages are those of January 1992, October 2002, and April

2011, and the corresponding samples end in 1991:Q3, 2002:Q2, and 2010:Q4. 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 stochastic-volatility estimates to those obtained from the structural-break specification. The latter generally captures the reduction in volatility after 1984, but not the recent increase during the Great Recession. Moreover, while the stochastic-volatility specification implies that volatility was gradually rising throughout the 1970s, the structural-break version is unable to capture this trend. Finally, the constant-volatility DSGE model systematically overstates volatility once the Great Moderation begins, because in significant part the model attempts to fit the high volatility before the Great Moderation.

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 credible bands (shaded area) of standard deviations of the structural shocks based on the DSGE model with SV-AR. The solid grey line is the posterior mean based on the model with a structural break in volatility (DV-SB). The dashed black line is the posterior mean based on the model with constant volatility.

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),

$$RMSE(i|h) = \sqrt{\frac{1}{P-h} \sum_{T=E}^{E+P-h} (Y_{i,T+h} - \hat{Y}_{i,T+h|T})^2},$$

where E is the starting point of the forecast evaluation sample (meaning it is the first forecast origin) and P is the number of forecast origins.

In Table 2 we present real-time forecast RMSE's for 1991:Q4 to 2011:Q1. We show RMSE's for the benchmark constant-volatility DSGE model in the first line of each panel, and RMSE ratios in the subsequent lines. Ratios less than one indicate that the forecasts from the corresponding time-varying volatility model are more accurate than the benchmark model forecasts. We use the following abbreviations: "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random walk stochastic-volatility process (14). In parentheses we show p-values of Diebold and Mariano (1995) tests of equal MSE against the one-sided alternative that the model with time-varying volatility is more accurate.

Table 2: Point Forecast RMSE's

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
		(a) Output Growth		
Const.	0.65	0.64	0.64	0.70
DV-SB	1.00 (0.45)	1.00 (0.57)	1.01 (0.85)	1.01 (0.90)
SV-AR	1.00 (0.64)	1.02 (1.00)	1.02 (0.99)	1.01 (0.95)
SV-RW	1.01 (0.88)	1.01 (1.00)	1.01 (1.00)	1.01 (1.00)
		(b) Inflation Rate		
Const.	0.24	0.27	0.30	0.37
DV-SB	0.99 (0.16)	1.00 (0.53)	0.99 (0.24)	0.93 (0.01)
SV-AR	1.06 (0.94)	1.01 (0.57)	0.94 (0.12)	0.85 (0.00)
SV-RW	1.00 (0.56)	0.99 (0.29)	0.94 (0.07)	0.86 (0.00)
		(c) Fed Funds Rate		
Const.	0.19	0.32	0.53	0.74
DV-SB	0.92 (0.00)	0.94 (0.01)	0.94 (0.01)	0.94 (0.01)
SV-AR	0.94 (0.00)	0.93 (0.00)	0.92 (0.00)	0.90 (0.00)
SV-RW	0.91 (0.00)	0.91 (0.00)	0.91 (0.00)	0.89 (0.00)

Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. 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 subsequent lines. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). In parentheses we show p-values of Diebold-Mariano tests of equal MSE against the one-sided alternative that the model with time-varying 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 $(P-h)^{1/3}$ in the other cases, where $P-h$ is the number of forecasting origins.

Forecasts from the specifications with time-varying volatility are significantly more accurate for the federal funds rate at all horizons, and for inflation at longer horizons. In contrast, output growth forecast accuracy is very similar across models and horizons. There is no clear ranking across the three time-varying volatility specifications. The simple deterministic break model performs slightly better in some instances and slightly worse than the stochastic-volatility versions in other instances. This basic scenario – allowing for time-varying 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 time-varying 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 $\{Y_{i,T+h}^{(j)}\}_{j=1}^{n_{sim}}$.

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 (first row of each cell) 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.[8] 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 makes clear that the constant-volatility DSGE model forecasts for output growth and inflation at all horizons and for the federal funds rate at the one-quarter horizon 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 are around 87 percent. Based on the reported t-statistic p-values, all empirical departures from 70 percent nominal coverage are statistically significant.

The coverage of the intervals from the models with time-varying volatility, 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. On balance, the SV-RW version yields more accurate coverage rates than the SV-AR specification. The structural break specification also does quite well. Thus, accounting for the volatility reduction in 1984 due to the Great Moderation appears to be of first-order importance for interval forecasts.

[8] Results for 90% credible intervals are similar; see the Online Appendix for details.

Table 3: 70 Percent Interval Forecast Evaluation (Coverage & Length)

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
		(a) Output Growth		
Const.	0.89 (0.00)	0.90 (0.00)	0.91 (0.00)	0.89 (0.00)
	1.96	1.99	2.00	2.00
DV-SB	0.76 (0.25)	0.73 (0.61)	0.71 (0.91)	0.69 (0.85)
	1.17	1.19	1.18	1.18
SV-AR	0.78 (0.08)	0.79 (0.08)	0.77 (0.21)	0.83 (0.04)
	1.40	1.44	1.48	1.54
SV-RW	0.76 (0.25)	0.75 (0.34)	0.76 (0.28)	0.69 (0.86)
	1.35	1.38	1.38	1.37
		(b) Inflation Rate		
Const.	0.87 (0.00)	0.87 (0.00)	0.85 (0.00)	0.86 (0.08)
	2.93	3.30	3.54	3.80
DV-SB	0.74 (0.38)	0.70 (0.99)	0.65 (0.54)	0.57 (0.24)
	1.96	2.24	2.39	2.57
SV-AR	0.76 (0.25)	0.75 (0.41)	0.79 (0.12)	0.77 (0.42)
	2.42	2.62	2.83	3.13
SV-RW	0.76 (0.25)	0.77 (0.25)	0.76 (0.39)	0.66 (0.71)
	2.30	2.55	2.70	2.89
		(c) Fed Funds Rate		
Const.	0.86 (0.00)	0.68 (0.81)	0.63 (0.51)	0.54 (0.26)
	2.20	3.05	3.93	4.55
DV-SB	0.62 (0.13)	0.52 (0.11)	0.49 (0.07)	0.51 (0.18)
	1.20	1.70	2.23	2.62
SV-AR	0.80 (0.04)	0.66 (0.66)	0.52 (0.11)	0.49 (0.13)
	1.46	1.99	2.55	3.02
SV-RW	0.73 (0.54)	0.61 (0.30)	0.51 (0.05)	0.49 (0.13)
	1.32	1.82	2.36	2.74

Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). Top row of each cell: 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 $(P-h)^{1/3}$ in the other cases, where $P-h$ is the number of forecasting origins. Bottom row of each cell: we report the average lengths of prediction intervals.

6.1.2 Interval Length

Table 3 (second row of each cell) also shows average prediction interval lengths. Average lengths based on the time-varying volatility models are roughly 30 percent shorter than those from the constant-volatility model. Hence the time-varying volatility intervals dominate on *both* the coverage and length dimensions. The intervals from the stochastic-volatility specifications are generally wider than those from the structural break specification. This seems by and large consistent with Figure 1, which indicates the estimated volatility from the SV-AR version often exceeds the post-break estimate from the DV-SB specification.

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's likelihood-ratio tests for 70-percent 1-step-ahead interval forecasts, 1991:Q4 to 2011:Q1. We show separate and joint tests for correct coverage and independence. The coverage tests consistently find no flaws in the time-varying volatility DSGE intervals, while simultaneously consistently finding severe flaws in the constant-volatility DSGE intervals. In general, the random-walk stochastic-volatility specification SV-RW leads to higher p-values than the SV-AR specification. The structural break specification fairs slightly better than SV-RW for the inflation rate and worse for the federal funds rate.

Neither the time-varying volatility nor the constant-volatility DSGE interval forecasts perform consistently well in terms of the independence test. In particular, the p-values for the federal funds rate tests are all less than 5 percent. This is not unexpected, 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. Incorporating stochastic volatility can naturally fix mis-calibration problems, but there is no way for it to fix inadequate conditional-mean dynamics.

Finally, the joint test considers both correct coverage and independence. The highest p-values are obtained for output growth and inflation forecasts by the random-walk stochastic-volatility and the structural-break specifications. All of the federal funds rate forecasts fail the joint test.

Table 4: Christoffersen Likelihood-Ratio Tests

Volatility	Coverage	Independence	Joint
	(a) Output Growth		
Const.	15.1 (0.00)	3.50 (0.06)	18.9 (0.00)
DV-SB	1.23 (0.27)	0.62 (0.43)	2.42 (0.30)
SV-AR	2.66 (0.10)	0.26 (0.61)	3.41 (0.18)
SV-RW	1.23 (0.27)	0.04 (0.85)	1.83 (0.40)
	(b) Inflation Rate		
Const.	12.9 (0.00)	0.10 (0.76)	13.2 (0.00)
DV-SB	0.73 (0.40)	1.10 (0.29)	2.42 (0.30)
SV-AR	1.23 (0.27)	6.43 (0.01)	8.23 (0.02)
SV-RW	1.23 (0.27)	1.90 (0.17)	3.69 (0.16)
	(c) Fed Funds Rate		
Const.	10.8 (0.00)	4.10 (0.04)	15.22 (0.00)
DV-SB	2.54 (0.11)	33.5 (0.00)	37.98 (0.00)
SV-AR	3.60 (0.06)	9.11 (0.00)	13.17 (0.00)
SV-RW	0.36 (0.55)	13.6 (0.00)	16.61 (0.00)

Notes: We show results for 70-percent 1-step-ahead interval forecasts. The real-time forecast sample is 1991:Q4 to 2011:Q1. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). We show Christoffersen's individual asymptotic $\chi^2(1)$ tests for coverage and for independence, as well as his joint asymptotic $\chi^2(2)$ test, 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. (2016).[9] The predictive likelihood is

$$S_M(h) = \frac{1}{P-h} \sum_{T=E}^{E+P-h} \log p(\tilde{Y}_{T+h}|Y_{1:T}), \quad h = 1, 2, ..., H, \tag{19}$$

where E is the starting point of the forecast evaluation sample, P is the number of forecast origins, and h is the forecast horizon. We distinguish $\tilde{Y}$ from Y because we exclude the 10-year-ahead inflation expectations INF_t^{10y} from the predictive likelihood. M denotes marginal, as opposed to joint, predictive likelihood, which can be defined as

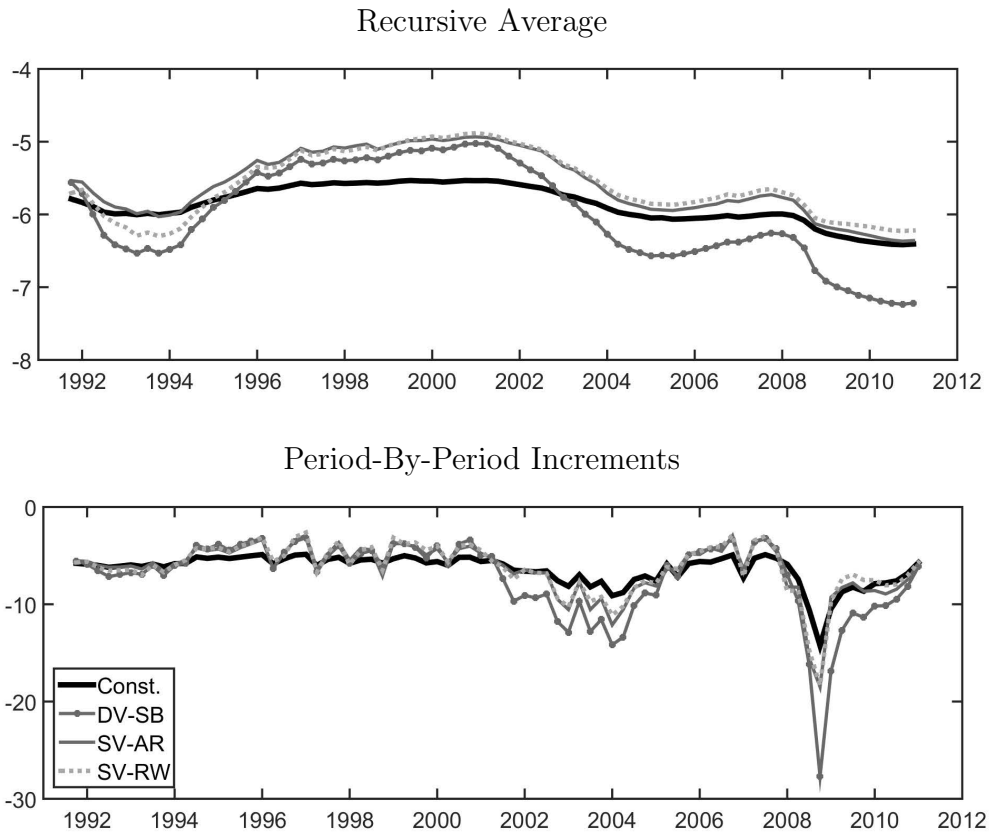
$$S_J(h) = \frac{1}{P-h} \sum_{T=E}^{E+P-h} \log p(\tilde{Y}_{T+1:T+h}|Y_{1:T}), \quad h = 1, 2, ..., H.$$

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 $p(Y_{T+h}|\nu_{T+1:T+h}^{(j)}, \theta^{(j)}, s_T^{(j)}, \nu_T^{(j)})$. 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

$$p(Y_{T+h}|Y_{1:T}) \approx \frac{1}{n_{sim}} \sum_{j=1}^{n_{sim}} p(Y_{T+h}|\nu_{T+1:T+h}^{(j)}, \theta^{(j)}, s_T^{(j)}, \nu_T^{(j)}). \tag{20}$$

[9] 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 1991:Q4 to 2011:Q1. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14).

In Figure 2 we show a time-series plot of 1-step-ahead predictive density values for GDP growth, inflation, and interest rates. The bottom panel shows period-by-period predictive likelihoods, whereas the top panel depicts the evolution of the recursive average $S_M(1)$ as the number of prediction periods P increases. Judging from the recursive average, after 1994 the two stochastic-volatility specifications dominate the constant volatility model. The structural break specification performs fairly well between 1994 and 2002, but then its performance starts to deteriorate.

The plot of the period-by-period predictive densities indicates that the relative ranking of the specifications changes over time. From 1994 to 2002 and 2005 to 2007, the time-varying volatility models perform better than the constant-volatility model. In the midst of the Great Recession 2008:Q4 output growth drops substantially and unexpectedly from the perspective of a wide variety of aggregate time series models. In this period the constant-volatility model fares better than its competitors because due to the pre-1984 observations, its estimated shock innovations are relatively large, in particular for the government spending and the technology shock (see Figure 1). Thus, the large drop in real activity appears less unexpected than for the time-varying volatility specifications. The stochastic-volatility models are able to adapt to the increase in macroeconomic volatility during the Great Recession (with a lag), whereas the structural break model is not.

Table 5: Marginal Predictive Likelihoods $S_M(h)$

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
(a) Output Growth				
Const.	-1.11	-1.11	-1.11	**-1.16**
DV-SB	**-0.99**	**-1.01**	**-1.03**	-1.17
SV-AR	-1.04	-1.12	-1.20	-1.62
SV-RW	-1.02	-1.08	-1.16	-1.49
(b) Inflation Rate				
Const.	-1.88	-1.91	-1.93	-2.05
DV-SB	-1.71	-1.74	-1.77	-1.92
SV-AR	-1.63	-1.66	**-1.70**	**-1.83**
SV-RW	**-1.62**	**-1.64**	-1.71	-1.88
(c) Fed Funds Rate				
Const.	-2.74	-2.85	-2.98	**-3.12**
DV-SB	-3.37	-3.61	-3.95	-4.46
SV-AR	-2.45	-2.57	**-2.93**	-3.92
SV-RW	**-2.37**	**-2.52**	-2.96	-4.26
(d) Multivariate				
Const.	-6.41	**-6.59**	**-6.81**	**-7.06**
DV-SB	-7.22	-7.63	-8.27	-9.26
SV-AR	-6.36	-6.70	-7.70	-11.46
SV-RW	**-6.22**	-6.61	-7.87	-12.87

Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We evaluate the predictive densities at the actuals obtained from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). 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.

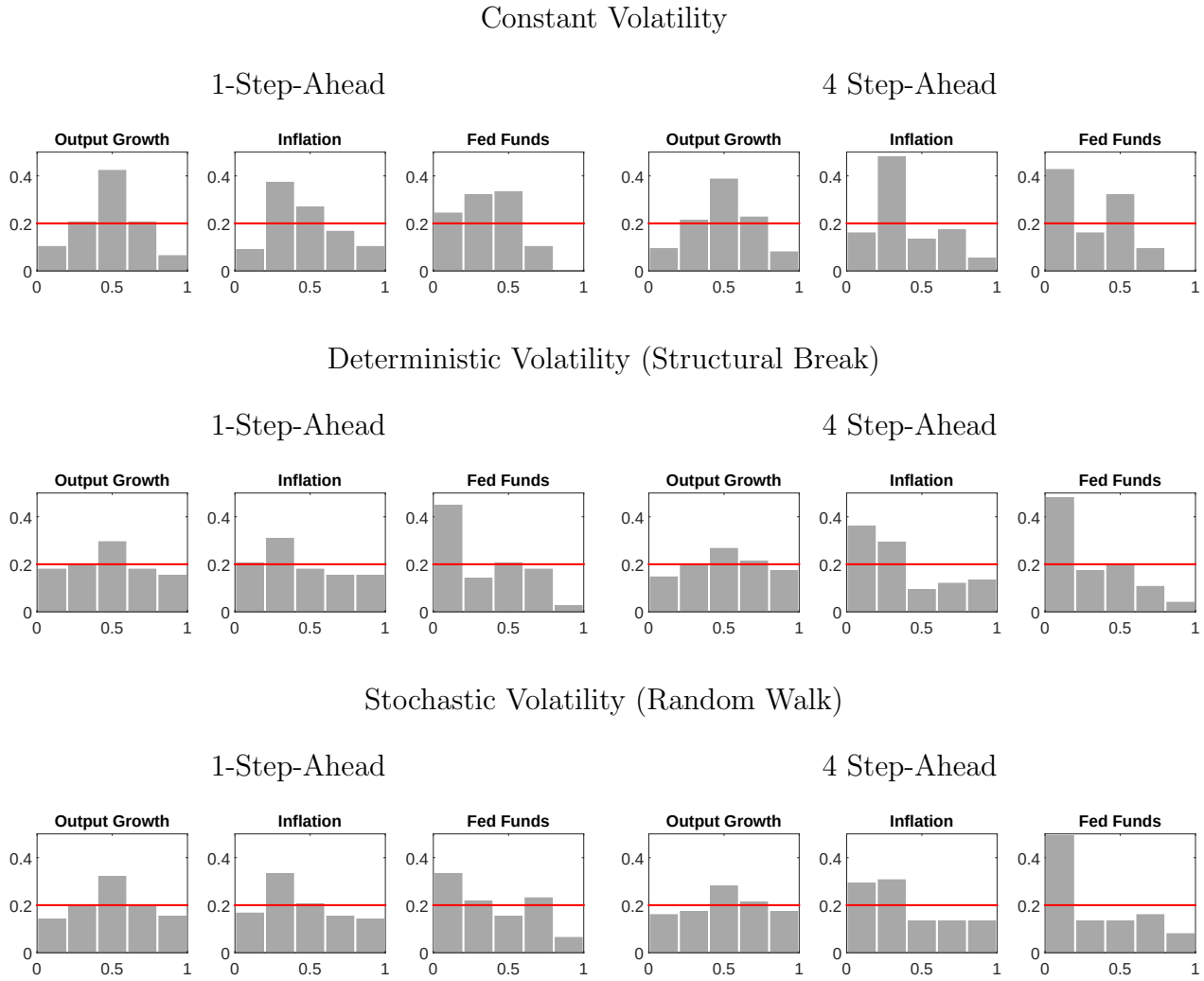
In Table 5 we present marginal predictive likelihoods for density forecasts at horizons $h = 1, 2, 4, 8$. From a univariate prediction perspective, the time-varying volatility specifications dominate the constant-volatility specification at horizons $h = 1, 2, 4$. The structural break specification fares slightly better for output growth, whereas the stochastic-volatility specifications work better for the inflation and federal funds rates. For $h = 1, 2$ the random walk stochastic-volatility process is preferred whereas for longer horizons the mean-reverting autoregressive process generates more accurate density forecasts. From a multivariate perspective, the random-walk stochastic volatility model is preferred at the one-step horizon (see also Figure 2). At $h = 2$ it is essentially a tie between the SV-RW and the constant-volatility specifications[10], whereas at horizons $h = 4, 8$ the constant-volatility version comes out ahead.

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.

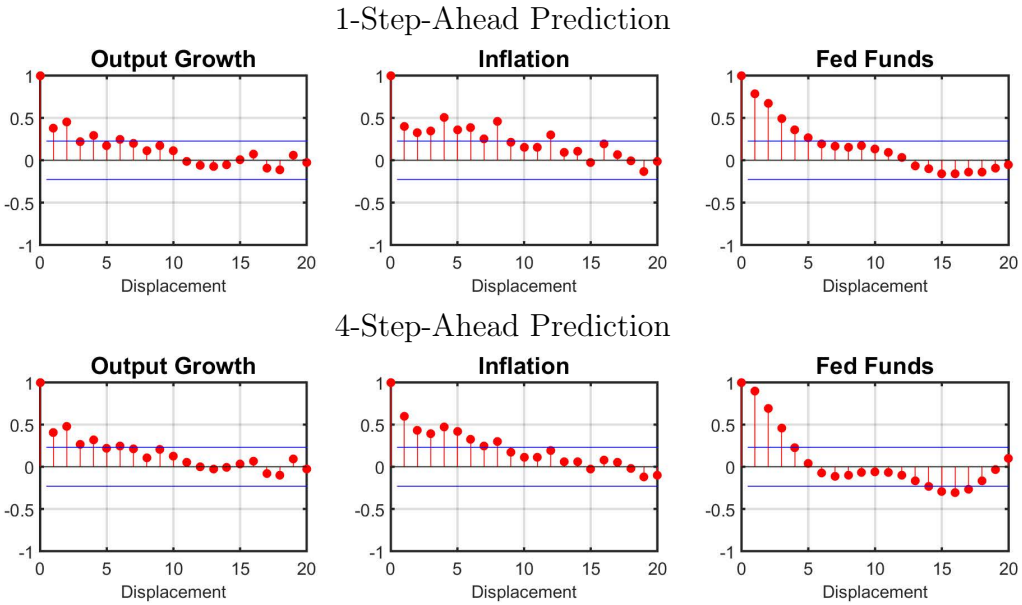
[10]See also the results for the evaluation sample that ends in 2007:Q4 reported in the Online Appendix

Figure 3: PIT Histograms



Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We calculate forecast errors using actuals obtained from the most recent vintage. 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 For Random Walk Stochastic Volatility



Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We calculate forecast errors using actuals obtained from the most recent vintage.

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,

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

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 constant and time-varying 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 structural break and the stochastic-volatility (random walk) specifications appear much more uniform.[11]

We present PIT sample autocorrelations in Figure 4. They essentially look the same for all four specifications considered in this paper. For brevity we only plot them for the random-

[11]The histogram for the mean-reverting stochastic-volatility processes looks similar to the one for the random-walk process.

walk stochastic-volatility model. Clear deviations from independence are apparent. Hence, although the time-varying-volatility DSGE models appear unconditionally well calibrated (in contrast to the constant-volatility model), 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. We also compared the performance of the stochastic-volatility specifications to a model with a one-time break in shock standard deviations at the end of 1984. The structural-break version performs generally better than the constant-volatility model, but is unable to adapt to the change in macroeconomic volatility during the Great Recession. 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.

Acknowledgments: For invaluable guidance we are grateful to the co-editors (Serge Darolles, Alain Monfort, and Eric Renault), and to two anonymous referees. For helpful comments we thank Fabio Canova, as well as participants at the Annual Conference on Real-Time Data Analysis, Methods, and Applications in Macroeconomics and Finance, the Federal Reserve

Bank of Philadelphia, the 2015 NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics, the 2015 NBER Summer Institute, the University of Pennsylvania, and European University Institute. For research support we thank the National Science Foundation (SES-1424843) and the Real-Time Data Research Center at the Federal Reserve Bank of Philadelphia.

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*, 35, 1659–1670.

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.

Curdia, Vasco, Marco Del Negro, and Daniel Greenwald (2014), "Rare Large Shocks and Great Recessions," *Journal of Applied Econometrics*, 29, 1031–1052.

Del Negro, M. and G.E. Primiceri (2015), "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum," *Review of Economic Studies*, 82, 1342–1345.

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.

Diebold, F.X. and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

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, 1935–1690.

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 (2016), "Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGE-VAR and VAR Models," *Journal of Applied Econometrics*, forthcoming.

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

Online Appendix

Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility

Francis X. Diebold, Frank Schorfheide, and Minchul Shin

A Evaluation of 90% Probability Interval Forecasts

We repeat the evaluation of interval forecasts presented in Section 6 for 90% probability intervals. The results are summarized in Tables A-1 and A-2. Including stochastic volatility generally improves the actual coverage rate of 90% predictive intervals.

Table A-1: 90 Percent Interval Forecast Evaluation (Coverage & Length)

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
		(a) Output Growth		
Const.	0.96 (0.01)	0.96 (0.07)	0.96 (0.08)	0.94 (0.39)
	3.12	3.17	3.18	3.18
DV-SB	0.89 (0.67)	0.87 (0.55)	0.87 (0.46)	0.86 (0.50)
	1.88	1.90	1.89	1.88
SV-AR	0.92 (0.44)	0.94 (0.33)	0.93 (0.37)	0.94 (0.39)
	2.38	2.46	2.56	2.68
SV-RW	0.91 (0.75)	0.90 (0.93)	0.91 (0.87)	0.89 (0.82)
	2.20	2.24	2.26	2.27
		(b) Inflation Rate		
Const.	0.96 (0.01)	0.99 (0.00)	0.99 (0.00)	0.94 (0.35)
	4.66	5.25	5.64	6.04
DV-SB	0.90 (0.94)	0.90 (0.91)	0.91 (0.87)	0.91 (0.81)
	3.13	3.57	3.82	4.09
SV-AR	0.92 (0.44)	0.96 (0.02)	0.97 (0.00)	0.94 (0.35)
	4.09	4.44	4.83	5.38
SV-RW	0.94 (0.20)	0.94 (0.28)	0.93 (0.37)	0.94 (0.35)
	3.83	4.26	4.57	4.98
		(c) Fed Funds Rate		
Const.	0.96 (0.01)	0.91 (0.86)	0.79 (0.16)	0.66 (0.07)
	3.50	4.85	6.26	7.27
DV-SB	0.87 (0.46)	0.69 (0.02)	0.59 (0.01)	0.54 (0.01)
	1.92	2.71	3.57	4.19
SV-AR	0.94 (0.20)	0.90 (0.94)	0.84 (0.44)	0.74 (0.18)
	2.54	3.48	4.52	5.42
SV-RW	0.94 (0.20)	0.87 (0.64)	0.77 (0.12)	0.69 (0.08)
	2.23	3.10	4.07	4.81

Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). Top row of each cell: we report the frequencies with which outcomes fall in 90-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 90 percent), calculated using Newey-West standard errors with bandwidth 0 at the 1-quarter horizon and $(P-h)^{1/3}$ in the other cases, where $P-h$ is the number of forecasting origins. Bottom row of each cell: we report the average lengths of prediction intervals.

Table A-2: Christoffersen Tests

Volatility	Coverage	Independence	Joint
	(a) Output Growth		
Const.	4.19 (0.04)	10.9 (0.00)	15.2 (0.00)
DV-SB	0.20 (0.66)	0.92 (0.34)	1.37 (0.51)
SV-AR	0.50 (0.48)	3.72 (0.05)	4.38 (0.11)
SV-RW	0.09 (0.76)	2.51 (0.11)	2.80 (0.25)
	(b) Inflation Rate		
Const.	4.19 (0.04)	3.14 (0.08)	7.41 (0.03)
DV-SB	0.01 (0.94)	0.04 (0.84)	0.26 (0.88)
SV-AR	0.50 (0.48)	0.57 (0.45)	1.22 (0.54)
SV-RW	1.26 (0.26)	1.11 (0.29)	2.51 (0.29)
	(c) Fed Funds Rate		
Const.	4.19 (0.04)	0.24 (0.62)	4.51 (0.11)
DV-SB	0.64 (0.42)	5.60 (0.02)	6.52 (0.04)
SV-AR	1.26 (0.26)	1.11 (0.29)	2.51 (0.29)
SV-RW	1.26 (0.26)	1.11 (0.29)	2.51 (0.29)

Notes: We show results for 90-percent 1-step-ahead interval forecasts. The real-time forecast sample is 1991:Q4 to 2011:Q1. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). We show tests for coverage and for independence, as well as joint tests, with p-values in parentheses.

B Ending the Evaluation Period in 2007:Q4

We now only use vintages up until January 2008. This last vintage utilizes NIPA data ending in 2007:Q3 and the 2007:Q4 observation is the last observation that we forecast.

Table A-3 summarizes the RMSEs for point forecasts. Compared to Table 2 in the main text the absolute values of the RMSEs are lower if the Great Recession is excluded from the forecast evaluation. The ranking of constant-volatility and time-varying volatility forecasts is largely unaffected.

Table A-4 shows the coverage probabilities. The p-values for output growth and inflation slightly drop, whereas the p-values for the federal funds rate intervals slightly increase. It remains true that the coverage probability of the models with time-varying volatility is generally closer to the nominal coverage probability of the predictive interval.

Table A-5 shows the results for the Christoffersen test. With respect to independence, the p-values for output growth increase whereas the p-values for inflation slightly decrease. On balance, the table leads to the same conclusions as the results reported in the main text.

Table A-6 shows average marginal predictive likelihoods. The pattern of highest marginal likelihood specifications is identical to the pattern reported in the main text.

Figure A-1 compares the distribution of one-step-ahead PITs for the samples that end in 2011:Q1 and 2007:Q4, respectively. For the shorter sample it is still true that allowing for time-varying volatility makes the PIT distribution more uniform. Moreover, the spike in the 0-0.2 for the interest rate forecasts is smoothed out, because there are now fewer periods in which the interest rate is zero.

Table A-3: Point Forecast RMSE's

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
		(a) Output Growth		
Const.	0.49	0.50	0.51	0.53
DV-SB	1.00 (0.55)	1.00 (0.57)	1.00 (0.69)	1.00 (0.68)
SV-AR	1.02 (0.93)	1.02 (0.99)	1.01 (0.99)	1.01 (0.98)
SV-RW	1.01 (0.97)	1.01 (1.00)	1.02 (1.00)	1.01 (0.99)
		(b) Inflation Rate		
Const.	0.21	0.24	0.26	0.29
DV-SB	0.99 (0.22)	1.00 (0.64)	1.00 (0.49)	0.93 (0.06)
SV-AR	1.13 (1.00)	1.05 (0.82)	0.98 (0.34)	0.81 (0.02)
SV-RW	1.02 (0.95)	1.00 (0.54)	0.96 (0.21)	0.82 (0.01)
		(c) Fed Funds Rate		
Const.	0.16	0.27	0.42	0.58
DV-SB	0.95 (0.00)	0.97 (0.06)	0.97 (0.05)	0.94 (0.03)
SV-AR	0.92 (0.00)	0.92 (0.01)	0.91 (0.01)	0.87 (0.01)
SV-RW	0.91 (0.00)	0.91 (0.01)	0.91 (0.01)	0.87 (0.02)

Notes: The real-time forecast sample is 1991:Q4 to 2007:Q4. 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 subsequent lines. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). In parentheses we show p-values of Diebold-Mariano tests of equal MSE against the one-sided alternative that the model with time-varying 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 $(P-h)^{1/3}$ in the other cases, where $P-h$ is the number of forecasting origins.

Table A-4: 70 Percent Interval Forecast Evaluation (Coverage & Length)

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
		(a) Output Growth		
Const.	0.92 (0.00)	0.94 (0.00)	0.94 (0.00)	0.93 (0.00)
	1.97	2.00	2.01	2.02
DV-SB	0.78 (0.10)	0.75 (0.32)	0.73 (0.64)	0.69 (0.90)
	1.15	1.17	1.17	1.17
SV-AR	0.80 (0.04)	0.83 (0.00)	0.81 (0.04)	0.86 (0.01)
	1.34	1.39	1.46	1.58
SV-RW	0.78 (0.10)	0.78 (0.12)	0.79 (0.07)	0.72 (0.79)
	1.29	1.32	1.34	1.39
		(b) Inflation Rate		
Const.	0.91 (0.00)	0.91 (0.00)	0.92 (0.00)	0.97 (0.00)
	2.94	3.33	3.59	3.87
DV-SB	0.78 (0.10)	0.75 (0.45)	0.69 (0.93)	0.66 (0.68)
	1.93	2.21	2.39	2.58
SV-AR	0.77 (0.19)	0.77 (0.33)	0.81 (0.06)	0.90 (0.00)
	2.35	2.56	2.78	3.12
SV-RW	0.77 (0.19)	0.77 (0.31)	0.77 (0.32)	0.72 (0.83)
	2.15	2.41	2.58	2.84
		(c) Fed Funds Rate		
Const.	0.91 (0.00)	0.80 (0.28)	0.76 (0.55)	0.66 (0.76)
	2.23	3.11	4.01	4.66
DV-SB	0.69 (0.89)	0.63 (0.48)	0.60 (0.35)	0.62 (0.60)
	1.20	1.70	2.25	2.67
SV-AR	0.83 (0.00)	0.73 (0.67)	0.63 (0.50)	0.59 (0.46)
	1.34	1.84	2.43	3.04
SV-RW	0.75 (0.31)	0.66 (0.63)	0.58 (0.25)	0.59 (0.46)
	1.19	1.65	2.18	2.69

Notes: The real-time forecast sample is 1991:Q4 to 2007:Q4. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). Top row of each cell: 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 $(P-h)^{1/3}$ in the other cases, where $P-h$ is the number of forecasting origins. Bottom row of each cell: we report the average lengths of prediction intervals.

Table A-5: Christoffersen Tests

Volatility	Coverage	Independence	Joint
	(a) Output Growth		
Const.	19.59 (0.00)	0.85 (0.36)	20.60 (0.00)
DV-SB	2.36 (0.12)	0.00 (0.96)	2.85 (0.24)
SV-AR	3.35 (0.07)	1.90 (0.17)	5.70 (0.06)
SV-RW	2.36 (0.12)	0.65 (0.42)	3.50 (0.17)
	(b) Inflation Rate		
Const.	16.52 (0.00)	1.24 (0.26)	17.95 (0.00)
DV-SB	2.36 (0.12)	2.39 (0.12)	5.24 (0.07)
SV-AR	1.56 (0.21)	5.36 (0.02)	7.44 (0.02)
SV-RW	1.56 (0.21)	3.44 (0.06)	5.53 (0.06)
	(c) Fed Funds Rate		
Const.	16.52 (0.00)	3.08 (0.08)	19.79 (0.00)
DV-SB	0.02 (0.89)	28.55 (0.00)	30.96 (0.00)
SV-AR	5.90 (0.02)	10.45 (0.00)	16.73 (0.00)
SV-RW	0.93 (0.33)	16.45 (0.00)	20.23 (0.00)

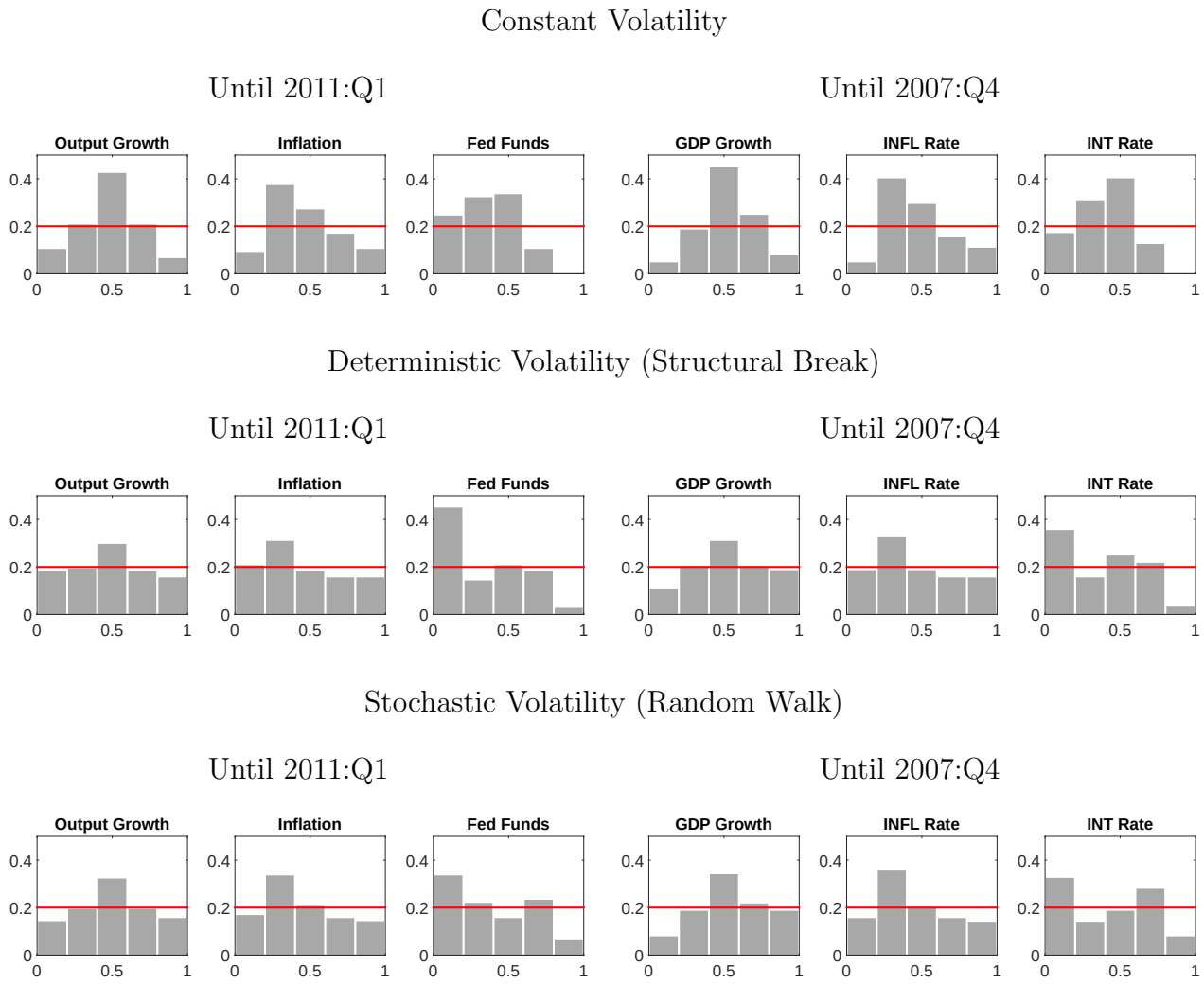
Notes: We show results for 70-percent 1-step-ahead interval forecasts. The real-time forecast sample is 1991:Q4 to 2007:Q4. We obtain "actuals" from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). We show tests for coverage and for independence, as well as joint tests, with p-values in parentheses.

Table A-6: Marginal Predictive Likelihoods $S_M(h)$

Volatility	$h = 1Q$	$h = 2Q$	$h = 4Q$	$h = 8Q$
(a) Output Growth				
Const.	-1.02	-1.02	-1.03	-1.07
DV-SB	**-0.73**	**-0.74**	**-0.76**	**-0.88**
SV-AR	-0.85	-0.86	-0.90	-1.03
SV-RW	-0.82	-0.84	-0.86	-0.98
(b) Inflation Rate				
Const.	-1.86	-1.87	-1.90	-2.01
DV-SB	-1.63	-1.65	-1.68	-1.84
SV-AR	-1.56	-1.58	**-1.62**	**-1.72**
SV-RW	**-1.54**	**-1.57**	**-1.62**	-1.73
(c) Fed Funds Rate				
Const.	-2.39	-2.45	-2.52	**-2.60**
DV-SB	-2.71	-2.83	-3.01	-3.27
SV-AR	-2.07	-2.11	**-2.29**	-2.62
SV-RW	**-2.03**	**-2.10**	-2.34	-2.82
(d) Multivariate				
Const.	-5.91	-6.02	**-6.18**	**-6.32**
DV-SB	-6.14	-6.36	-6.75	-7.29
SV-AR	-5.67	**-5.80**	-6.36	-7.52
SV-RW	**-5.61**	**-5.80**	-6.55	-8.11

Notes: The real-time forecast sample is 1991:Q4 to 2007:Q4. We evaluate the predictive densities at the actuals obtained from the most recent vintage. "Const." is constant volatility; "DV-SB" is deterministic volatility with structural break (17); "SV-AR" is AR(1) stochastic-volatility process (16); and "SV-RW" is random-walk stochastic-volatility process (14). 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.

Figure A-1: PIT Histograms



Notes: The real-time forecast sample is 1991:Q4 to 2011:Q1. We calculate forecast errors using actuals obtained from the most recent vintage. 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.