ISSN 1936-5330

The Federal Reserve Bank *of* Kansas City
Economic Research Department

Real-Time Density Forecasts from VARs with Stochastic Volatility

Todd E. Clark

June 2009; Revised July 2010
RWP 09-08

Research Working Papers

Real-Time Density Forecasts from VARs with Stochastic Volatility

Todd E. Clark*

First Version: June 2009
This version: July 2010

RWP 09-08

Abstract

Central banks and other forecasters are increasingly interested in various aspects of density forecasts. However, recent sharp changes in macroeconomic volatility – such as the Great Moderation and the more recent sharp rise in volatility associated with greater variation in energy prices and the deep global recession – pose significant challenges to density forecasting. Accordingly, this paper examines, with real-time data, density forecasts of U.S. GDP growth, unemployment, inflation, and the federal funds rate from BVAR models with stochastic volatility. The results indicate that adding stochastic volatility to BVARs materially improves the real-time accuracy of density forecasts.

Keywords: Steady-state prior, Prediction, Bayesian methods

JEL Classification: C53, C32, E37

*Economic Research Dept.; Federal Reserve Bank of Kansas City; 1 Memorial Drive; Kansas City, MO 64198; todd.e.clark@kc.frb.org; phone: (816)881-2575; fax: (816)881-2199. The author gratefully acknowledges helpful conversations with Taeyoung Doh, John Geweke, Mattias Villani, and Chuck Whiteman and comments from anonymous referees, Gianni Amisano, Marek Jarocinski, Michael McCracken, Par Osterholm, Shaun Vahey, seminar participants at the Federal Reserve Banks of Boston, Chicago, and Kansas City, and participants at the 2010 ECB-Bundesbank-EABCN workshop on forecasting techniques. Special thanks are due to Mattias Villani for providing simplified forms of GLS-based equations for the slope and steady state coefficients of the VAR with a steady state prior. The views expressed herein are solely those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

1 Introduction

Policymakers and forecasters are increasingly interested in forecast metrics that require density forecasts of macroeconomic variables. Such metrics include confidence intervals, fan charts, and probabilities of recession or inflation exceeding or falling short of a certain threshold. For example, in 2008 the Federal Reserve expanded its publication of forecast information to include qualitative indications of the uncertainty surrounding the outlook. Other central banks, such as the Bank of Canada, Bank of England, Norges Bank, South African Reserve Bank, and Sveriges Riksbank, routinely publish fan charts that provide entire forecast distributions for inflation and, in some nations, a measure of output or the policy interest rate.

For many countries, however, changes in volatility over time pose a challenge to density forecasting. The Great Moderation significantly reduced the volatility of many macroeconomic variables. More recently, though, a variety of forces have substantially increased volatility (see, e.g., Clark (2009)). In the few years before the 2007-2009 recession, increased volatility of energy prices caused the volatility of total inflation to rise sharply. Then, the severe recession raised the volatility of a range of macroeconomic variables — by enough to largely (although probably temporarily) reverse the Great Moderation in GDP growth.

Such shifts in volatility have the potential to result in forecast densities that are either far too wide or too narrow. For example, until recently the volatility of U.S. growth and inflation was much lower in data since the mid-1980s than in data for the 1970s and early 1980s. Density forecasts for GDP growth in 2007 based on time series models assuming constant variances over a sample such as 1960-2006 would probably be far too wide. On the other hand, in late 2008, density forecasts for 2009 based on time series models assuming constant variances for 1985-2008 would probably be too narrow. Results in Jore, Mitchell, and Vahey (2010) support this intuition. In an analysis of real-time density forecasts since the mid-1980s, they find that models estimated with full samples of data and constant parameters fare poorly in density forecasting. Allowing discrete breaks in variances materially improves density forecasts made in the Great Moderation period.

If volatility breaks were rare and always observed clearly with hindsight, simple split-sample or rolling sample methods might be used to obtain reliable density forecasts. But as recent events have highlighted, breaks such as the Great Moderation once thought to be effectively permanent can turn out to be shorter-lived, and reversed (at least temporarily).

Over time, then, obtaining reliable density forecasts likely requires forecast methods that allow for repeated breaks in volatilities.

Accordingly, this paper examines the accuracy of real-time density forecasts of U.S. macroeconomic variables made with Bayesian vector autoregressions (BVARs) that allow for continuous changes in the conditional variances of the model's shocks — that is, stochastic volatility, as in such studies as Cogley and Sargent (2005) and Primiceri (2005). The forecasted variables consist of GDP growth, unemployment, inflation, and the federal funds rate. While many studies have examined point forecasts from VARs in similar sets of variables, density forecasts have received much less attention. Cogley, Morozov, and Sargent (2005) and Beechey and Osterholm (2008) present density forecasts from BVARs estimated for the U.K. and Australia, but only for a single point in time, rather than a longer period of time that would allow historical evaluation. While Jore, Mitchell, and Vahey (2010) provide an historical evaluation of density forecasts, their volatility models are limited to discrete break specifications. I extend this prior work by examining the historical accuracy of density forecasts from BVARs with a general volatility model — specifically, stochastic volatility.

In light of the evidence in Clark and McCracken (2008, 2010) that the accuracy of point forecasts of GDP growth, inflation, and interest rates is improved by specifying the inflation and interest rates as deviations from trend inflation, the model of interest in this paper also specifies the unemployment rate, inflation, and interest rate variables in gap, or deviation from trend, form. In addition, based on a growing body of evidence on the accuracy of point forecasts, the BVAR of interest incorporates an informative prior on the steady state values of the model variables. Villani (2009) develops a Bayesian estimator of a (constant variance) VAR with an informative prior on the steady state. Applications of the estimator in studies such as Adolfson, et al. (2007), Beechey and Osterholm (2008), Osterholm (2008), and Wright (2010) have shown that the use of a prior on the steady state often improves the accuracy of point forecasts. In a methodological sense, this paper extends the estimator of Villani (2009) to include stochastic volatility.

The evidence presented in the paper shows that adding stochastic volatility to the BVAR with most variables in gap form and a steady state prior materially improves real-time density forecasts. Compared to models with constant variances, models with stochastic volatility have significantly more accurate interval forecasts (coverage rates), normalized forecast errors (computed from the probability integral transforms, or PITs) that are much closer to a standard normal distribution, and average log predictive density scores that are much lower. Adding stochastic volatility to univariate AR models also materially improves density forecast calibration relative to AR models with constant variances. In the case of BVARs, adding stochastic volatility also improves the accuracy of point forecasts, lowering root mean square errors (RMSEs).

Section 2 describes the real-time data used. Section 3 presents the BVAR with stochastic volatility and an informative prior on the steady state means. Section 4 details the other forecasting models considered. Section 5 reports the results. Section 6 concludes.

2 Data

Forecasts are evaluated for four variables: output growth, the unemployment rate, inflation, and the federal funds rate. As detailed in Section 3, the primary BVAR specification of interest also includes as an endogenous variable the long-term inflation expectation from the Blue Chip Consensus, which is used to measure trend inflation. As detailed in section 3, the survey expectation is included to account for uncertainty associated with the inflation trend.

Output is measured as GDP or GNP, depending on data vintage. Inflation is measured with the GDP or GNP deflator or price index. Growth and inflation rates are measured as annualized log changes (from $t-1$ to t). Quarterly real-time data on GDP or GNP and the GDP or GNP price series are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). For simplicity, hereafter "GDP" and "GDP price index" refer to the output and price series, even though the measures are based on GNP and a fixed weight deflator for much of the sample. In the case of unemployment and fed funds rates, for which real-time revisions are small to essentially non–existent, I simply abstract from real-time aspects of the data. The quarterly data on unemployment and the interest rate are constructed as simple within-quarter averages of the source monthly data (in keeping with the practice of, e.g., Blue Chip and the Federal Reserve).

The long-term inflation expectation is measured as the Blue Chip Consensus forecast of average GDP price inflation 6-10 years ahead. The Blue Chip forecasts are taken from surveys published in the spring and fall of each year from 1979 through 2008. For model estimation purposes, the Blue Chip data are extended from 1979 back to 1960 with an estimate of expected GDP inflation based on exponential smoothing (with a smoothing parameter of 0.05). As noted by Kozicki and Tinsley (2001a,b) and Clark and McCracken (2008), exponential smoothing yields an estimate that matches up reasonably well with survey–based measures of long–run expectations in data since the early 1980s. A not-for-publication appendix provides additional detail on the real-time series of inflation expectations.

The full forecast evaluation period runs from 1985:Q1 through 2008:Q3, which involves real-time data vintages from 1985:Q1 through 2009:Q1. As described in Croushore and Stark (2001), the vintages of the RTDSM are dated to reflect the information available around the middle of each quarter. Normally, in a given vintage t, the available NIPA data run through period $t-1$. For each forecast origin t starting with 1985:Q1, I use the real-time data vintage t to estimate the forecast models and construct forecasts for periods t and beyond. For forecasting models estimated recursively (see section 3.4), the starting point of the model estimation sample is always 1961:Q1.

The results on forecast accuracy cover forecast horizons of 1 quarter ($h = 1Q$), 2 quarters ($h = 2Q$), 1 year ($h = 1Y$), and 2 years ($h = 2Y$) ahead. In light of the time $t-1$ information actually incorporated in the VARs used for forecasting at t, the 1-quarter ahead forecast is a current quarter (t) forecast, while the 2-quarter ahead forecast is a next quarter ($t+1$) forecast. In keeping with Federal Reserve practice, the 1– and 2–year ahead forecasts for GDP growth and inflation are 4–quarter rates of change (the 1–year ahead forecast is the percent change from period t through $t+3$; the 2–year ahead forecast is the percent change from period $t+4$ through $t+7$). The 1– and 2–year ahead forecasts for unemployment and the funds rate are quarterly levels in periods $t+3$ and $t+7$, respectively.

As discussed in such sources as Romer and Romer (2000), Sims (2002), and Croushore (2005), evaluating the accuracy of real-time forecasts requires a difficult decision on what to take as the actual data in calculating forecast errors. The GDP data available today for, say, 1985, represent the best available estimates of output in 1985. However, output as defined and measured today is quite different from output as defined and measured in 1970. For example, today we have available chain-weighted GDP; in the 1980s, output was measured with fixed-weight GNP. Forecasters in 1985 could not have foreseen such changes and the potential impact on measured output. Accordingly, I follow studies such as Romer and Romer (2000) and Faust and Wright (2009) and use the second available estimates of GDP/GNP and the GDP/GNP deflator as actuals in evaluating forecast accuracy. In the case of h–step ahead (for h = 1Q, 2Q, 1Y, and 2Y) forecasts made for period $t + h$ with vintage t data ending in period $t - 1$, the second available estimate is normally taken from the vintage $t + h + 2$ data set. In light of my abstraction from real-time revisions in unemployment and the funds rate, for these series the real-time data correspond to the final vintage data.

3 BVAR with stochastic volatility and informative priors on steady state means (BVAR-SSPSV)

The model of primary interest, denoted BVAR-SSPSV — short for BVAR with most variables in gap form, an informative steady state prior, and stochastic volatility — extends Villani's (2009) model with a steady state prior to include stochastic volatility, modeled as in Cogley and Sargent (2005). As noted in the introduction, the use of gaps and steady state priors is motivated by prior research on the benefits to the accuracy of point forecasts. Stochastic volatility is added in the hope of improving density forecasts in the face of likely changes in shock variances. This section details the treatment of trends, the model, estimation procedure, priors, and the generation of posterior distributions of forecasts.

3.1 Trends

In the BVARs with steady state priors, the unemployment rate, inflation, and funds rate variables are specified in gap, or deviation from trend, form, with the trends measured in real time. The trend specifications are based in part on the need to be able to easily and tractably account for the impact of trend uncertainty on the forecast distributions. Unemployment u_t is centered around a trend u^*_{t-1} computed by exponential smoothing, with a smoothing coefficient of 0.02: $u^*_t = u^*_{t-1} + 0.02(u_t - u^*_{t-1})$. The smoothing coefficient setting of 0.02 suffices to yield a slow-moving trend; using a coefficient of 0.05 yields a more variable trend but very similar forecast results. As emphasized by Cogley (2002), exponential smoothing offers a simple and computationally convenient approach to capturing gradual changes in means. In general, exponential smoothing has also long been known to be effective for trend estimation and forecasting (e.g., Makridakis and Hibon (2000) and Chatfield, et al. (2001)). In this case, the use of exponential smoothing makes it easy to form trend unemployment forecasts over the forecast horizon, and thereby incorporate the effects of trend uncertainty in the forecast distributions for unemployment.

Inflation and the funds rate are centered around the long-term inflation expectation from

Blue Chip, described in Section 2. To account for the uncertainty in the forecasts of inflation and the funds rate associated with the trend defined as the long-run inflation expectation, the BVARs with steady state priors include the change in the expectation as an endogenous variable, which is forecast along with the other variables of the system. However, the inclusion of the long-run expectation as an endogenous variable does not appear to give the model with the steady state prior an advantage over the simple BVAR. A model without the expectation as an endogenous variable, in which the inflation expectation is assumed constant (at its last observed value) over the forecast horizon, yields results similar to those reported for the BVAR-SSP specifications that endogenize the expectation.

3.2 Model

Let y_t denote the $p \times 1$ vector of model variables and d_t denote a $q \times 1$ vector of deterministic variables. In this implementation, y_t includes GDP growth, the unemployment rate less its trend lagged one period, inflation less the long-run inflation expectation, the funds rate less the long-run inflation expectation, and the change in the long-run inflation expectation. In this paper, the only variable in d_t is a constant. Let $\Pi(L) = I_p - \Pi_1 L - \Pi_2 L^2 - \cdots - \Pi_k L^k$, $\Psi =$ a $p \times q$ matrix of coefficients on the deterministic variables, and $A =$ a lower triangular matrix with ones on the diagonal and coefficients a_{ij} in row i and column j (for $i = 2, \ldots, p$, $j = 1, \ldots, i-1$). The VAR(k) with stochastic volatility takes the form

$$\begin{aligned} \Pi(L)(y_t - \Psi d_t) &= v_t, \\ v_t &= A^{-1}\Lambda_t^{0.5}\epsilon_t,\ \epsilon_t \sim N(0, I_p),\ \Lambda_t = \text{diag}(\lambda_{1,t}, \lambda_{2,t}, \lambda_{3,t}, \ldots, \lambda_{p,t}) \qquad (1) \\ \log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t},\ \nu_{i,t} \sim \text{iid } N(0, \phi_i)\ \forall\ i = 1, \ldots, p. \end{aligned}$$

Under the stochastic volatility model, taken from Cogley and Sargent (2005), the log variances in Λ_t follow random walk processes. The (diagonal) variance-covariance matrix of the vector of innovations to the log variances is denoted Φ. This particular representation provides a simple and general approach to allowing time variation in the variances and covariances of the residuals v_t. Under the above specification, the residual variance–covariance for period t is $\text{var}(v_t) = \Sigma_t \equiv A^{-1}\Lambda_t A^{-1\prime}$.

3.3 Estimation procedure

The model is estimated with a five-step Metropolis-within-Gibbs MCMC algorithm, combining modified portions of the algorithms of Cogley and Sargent (2005) and Villani (2009).

[Special thanks are due to Mattias Villani for providing the formulae for posterior means and variances of Π and Ψ, which generalize the constant-variance formulae in Villani (2009).] The Metropolis step is used for the estimation of stochastic volatility, following Cogley and Sargent (2005) in their use of the Jacquier, Polson, and Rossi (1994) algorithm. If I instead used the algorithm of Kim, Shephard, and Chib (1998) for stochastic volatility estimation, the Metropolis step would be replaced with another Gibbs sampling step. However, in preliminary investigations with BVAR models, estimates based on the Kim, Shephard, and Chib algorithm seemed to be unduly dependent on priors and prone to yielding highly variable (across data samples) estimates of volatilities.

Step 1: Draw the slope coefficients Π conditional on Ψ, the history of Λ_t, A, and Φ.

For this step, the VAR is recast in demeaned form, using $Y_t = y_t - \Psi d_t$:

$$Y_t = (I_p \otimes X_t') \cdot \text{vec}(\Pi) + v_t, \ \text{var}(v_t) = \Sigma_t = A^{-1}\Lambda_t A^{-1\prime}, \tag{2}$$

where X_t contains the appropriate lags of Y_t and vec(Π) contains the VAR slope coefficients.

The vector of coefficients is sampled from a normal posterior distribution with mean $\bar{\mu}_\Pi$ and variance $\bar{\Omega}_\Pi$, based on prior mean μ_Π and Ω_Π, where:

$$\bar{\Omega}_\Pi^{-1} = \Omega_\Pi^{-1} + \sum_{t=1}^{T}(\Sigma_t^{-1} \otimes X_t X_t') \tag{3}$$

$$\bar{\mu}_\Pi = \bar{\Omega}_\Pi \left\{ \text{vec}(\sum_{t=1}^{T} \Sigma_t^{-1} Y_t X_t') + \Omega_\Pi^{-1}\mu_\Pi \right\}. \tag{4}$$

Step 2: Draw the steady state coefficients Ψ conditional on Π, the history of Λ_t, A, and Φ.

For this step, the VAR is rewritten as

$$q_t = \Pi(L)\Psi d_t + v_t, \ \text{where } q_t \equiv \Pi(L) y_t. \tag{5}$$

The dependent variable q_t is obtained by applying to the vector y_t the lag polynomial estimated with the preceding draw of the Π coefficients. The right-hand side term $\Pi(L)\Psi d_t$ simplifies to $\Theta \bar{d}_t$, where, as in Villani (2009) with some modifications, $\bar{d}_t$ contains current and lagged values of the elements of d_t, and Θ is defined such that $\text{vec}(\Theta) = U\text{vec}(\Psi)$:

$$\bar{d}_t = (d_t', -d_{t-1}', -d_{t-2}', -d_{t-3}', \ldots, -d_{t-k}')' \tag{6}$$

$$U = \begin{pmatrix} I_{pq \times pq} \\ I_q \otimes \Pi_1 \\ I_q \otimes \Pi_2 \\ I_q \otimes \Pi_3 \\ \vdots \\ I_q \otimes \Pi_k \end{pmatrix}. \tag{7}$$

The vector of coefficients Ψ is sampled from a normal posterior distribution with mean $\bar{\mu}_\Psi$ and variance $\bar{\Omega}_\Psi$, based on prior mean μ_Ψ and Ω_Ψ, where:

$$\bar{\Omega}_\Psi^{-1} = \Omega_\Psi^{-1} + U' \Big\{ \sum_{t=1}^{T} (\bar{d}_t \bar{d}_t' \otimes \Sigma_t^{-1}) \Big\} U \tag{8}$$

$$\bar{\mu}_\Psi = \bar{\Omega}_\Psi \Bigg\{ U' \text{vec}(\sum_{t=1}^{T} \Sigma_t^{-1} q_t \bar{d}_t') + \Omega_\Psi^{-1} \mu_\Psi \Bigg\}. \tag{9}$$

<u>Step 3</u>: Draw the elements of A conditional on Π, Ψ, the history of Λ_t, and Φ.

Following Cogley and Sargent (2005), rewrite the VAR as

$$A\Pi(L)(y_t - \Psi d_t) \equiv A\hat{y}_t = \Lambda_t^{0.5} \epsilon_t, \tag{10}$$

where, conditional on Π and Ψ, $\hat{y}_t$ is observable. This system simplifies to a set of $i = 2, \ldots, p$ equations, with equation i having as dependent variable $\hat{y}_{i,t}$ and as independent variables $-1 \cdot \hat{y}_{j,t}, j = 1, \ldots, i-1$, with coefficients a_{ij}. Multiplying equation i by $\lambda_{i,t}^{-0.5}$ eliminates the heteroskedasticity associated with stochastic volatility. Then, proceeding separately for each transformed equation i, draw the i'th equation's vector of j coefficients a_{ij} from a normal posterior distribution with the mean and variance implied by the posterior mean and variance computed in the usual (OLS) way. See Cogley and Sargent (2005) for details.

<u>Step 4</u>: Draw the elements of the variance matrix Λ_t conditional on Π, Ψ, A, and Φ.

Following Cogley and Sargent (2005), the VAR can be rewritten as

$$A\Pi(L)(y_t - \Psi d_t) \equiv \tilde{y}_t = \Lambda_t^{0.5} \epsilon_t, \tag{11}$$

where $\epsilon_t \sim N(0, I_p)$. Taking logs of the squares yields

$$\log \tilde{y}_{i,t}^2 = \log \lambda_{i,t} + \log \epsilon_{i,t}^2, \ \forall \ i = 1, \ldots, p. \tag{12}$$

The conditional volatility process is

$$\log(\lambda_{i,t}) = \log(\lambda_{i,t-1}) + \nu_{i,t}, \ \nu_{i,t} \sim \text{iid } N(0, \phi_i) \ \forall \ i = 1, \ldots, p. \tag{13}$$

The estimation of the time series of $\lambda_{i,t}$ proceeds equation by equation, using the measured $\log \tilde{y}_{i,t}^2$ and Cogley and Sargent's (2005) version of the Metropolis algorithm of Jacquier, Polson, and Rossi (1994).

Step 5: Draw the innovation variance matrix Φ conditional on Π, Ψ, the history of Λ_t, and A.

Following Cogley and Sargent (2005), the sampling of the diagonal elements of Φ, the variances of innovations to log volatilities, is based on inverse Wishart priors and posteriors. For each equation i, the posterior scaling matrix is a linear combination of the prior and the sample variance innovations computed as the variance of $\lambda_{i,t} - \lambda_{i,t-1}$. I obtain draws of each Φ_i by sampling from the inverse Wishart posterior with this scale matrix.

3.4 Priors and other estimation details

While the BVAR-SSPSV directly models variation over time in the means of most variables and in conditional variances, it is possible that the slope coefficients of the VAR could have drifted some over time. Accordingly, I consider forecasts from model estimates generated with both recursive (allowing the data sample to expand as forecasting moves forward in time) and rolling (keeping the estimation sample fixed at 80 observations and moving it forward as forecasting moves forward) schemes. The use of a 20-year rolling window follows such studies as Del Negro and Schorfheide (2004). The rolling scheme does not much affect stochastic volatility estimates, which are quite similar across the recursive and rolling specifications. It has a larger impact on estimates of VAR slope coefficients and steady states (Ψ), which in some cases differ quite a bit across the recursive and rolling specifications.

As to priors, the prior for the VAR slope coefficients $\Pi(L)$ is based on a Minnesota specification. The prior means suppose each variable follows an AR(1) process, with coefficients of 0.25 for GDP growth and 0.8 for the other variables. Prior standard deviations are controlled by the usual hyperparameters, with overall tightness of 0.2, cross–equation tightness of 0.5, and linear decay in the lags. The standard errors used in setting the prior are estimates from univariate AR(4) models fit with a training sample consisting of the 40 observations preceding the estimation sample used for a given vintage.

Priors are imposed on the deterministic coefficients Ψ to push the steady-states toward certain values, of: (1) GDP growth, 3.0 percent; (2) unemployment less the exponentially smoothed trend, 0.0; (3) inflation less the long-run inflation expectation of Blue Chip, 0.0; (4) federal funds rate less the long-run inflation expectation of Blue Chip, 2.5; and (5) change in the long-run inflation expectation of Blue Chip, 0.0. Accordingly, in the prior for the elements of Ψ, all means are zero, except as follows: GDP growth, intercept coefficient of 3.0; and fed funds rate, intercept coefficient of 2.5. In the recursive (rolling) estimation, I set the following standard deviations on each element of Ψ: GDP growth, 0.2 (0.3); unemployment less trend, 0.2 (0.3); inflation less long-run expectation, 0.2 (0.3); fed funds rate less long-run inflation expectation, 0.6 (0.75); and change in long-run inflation expectation, 0.2 (0.2). I use slightly tighter steady state priors for the recursive scheme than the rolling because, in the recursive case, the gradual increase in the size of the estimation sample (as forecasting moves forward) gradually reduces the influence of the prior.

For the volatility portion of the model, I use uninformative priors for the elements of A and loose priors for the initial values of $\log(\lambda_{i,t})$ and the variances of the innovations to $\log(\lambda_{i,t})$. The prior settings are similar to those used in other analyses of VARs with stochastic volatility (e.g., Cogley and Sargent (2005) and Primiceri (2005)), except that, in light of extant evidence of volatility changes, the prior mean on the variances of shocks to volatility is set in line with the higher value of Stock and Watson (2007) than the very low value used by Cogley and Sargent (2005) and Primiceri (2005). More specifically, I use the following priors:

$$\begin{aligned}
\log \lambda_{i,0} &\sim N(\log \hat{\lambda}_{i,OLS},\ 4)\ \forall\ i = 1, \ldots, p \\
a_i &\sim N(0, 1000^2 \cdot I_{i-1})\ \forall\ i = 2, \ldots, p \\
\phi_i &\sim IW(5 \cdot 0.035,\ 5)\ \forall\ i = 1, \ldots, p,
\end{aligned}$$

where a_i denotes the $(i-1) \times 1$ vector of $a_{i,j}$ coefficients in the i'th row of A and the $\hat{\lambda}_{i,OLS}$ are simple residual variances from AR(4) models estimated with a training sample of the 40 observations preceding the estimation sample. The variance of 4 on each $\log \lambda_{i,0}$ corresponds to a quite loose prior on the initial variances, in light of the log transformation of the variances.

3.5 Drawing forecasts

For each (retained) draw in the MCMC chain, I draw forecasts from the posterior distribution using an approach like that of Cogley, Morozov, and Sargent (2005). To incorporate uncertainty associated with time variation in Λ_t over the forecast horizon of 8 periods, I sample innovations to Λ_{t+h} from a normal distribution with (diagonal) variance Φ, and use the random walk specification to compute Λ_{t+h} from Λ_{t+h-1}. For each period of the forecast horizon, I then sample shocks to the VAR with a variance of Σ_{t+h} and compute the forecast draw of Y_{t+h} from the VAR structure and drawn shocks.

In all forecasts obtained from models with steady state priors, the model specification readily permits the construction of forecast distributions that account for the uncertainty associated with the trend unemployment rate and long-run inflation expectation. In each draw, the model is used to forecast GDP growth, unemployment less trend lagged one period, inflation less the long-run inflation expectation, the funds rate less the long-run inflation expectation, and the change in the long-run inflation expectation. The forecasted changes in the long-run expectation are accumulated and added to the value at the end of the estimation sample to obtain the forecasted level of the expectation. The forecasts of the level of the expectation are then added to the forecasts of inflation less the expectation and the funds rate less the expectation to obtain forecasts of the levels of inflation and the funds rate. Forecasts of the level of the unemployment rate and the exponentially smoothed trend are obtained by iterating forward, adding the lagged trend value to obtain the forecast of the unemployment rate, then computing the current value of the unemployment trend, and continuing forward in time over the forecast horizon.

Finally, I report posterior estimates based on 10,000 draws, obtained by first generating 10,000 burn-in draws and then saving every fifth draw from another 50,000 draws. Point forecasts are constructed as posterior means of the MCMC distributions. In most cases, the forecasts and forecast errors pass simple normality tests, supporting the use of means.

4 Other Models Considered

To establish the effectiveness of steady state priors and stochastic volatility, forecasts from the BVAR-SSPSV model are compared against a range of forecasts from other models. Because point forecasts from VARs are often dominated (post-1984) by point forecasts from univariate models (see, e.g., Clark and McCracken (2008, 2010)), the set of models includes AR models with constant error variances and with stochastic volatility. The set of models also includes conventional BVARs without steady state priors or stochastic volatility and BVARs with steady priors and not stochastic volatility.

4.1 AR models

The set of univariate models is guided by prior evidence (e.g., Clark and McCracken (2008, 2010) and Stock and Watson (2007)) on the accuracy of point forecasts and by the practical need for specifications that readily permit (i) constant variances and stochastic volatility and (ii) estimation by MCMC methods (for comparability, the same ones I use for the BVARs) for the purpose of obtaining forecast densities. For output growth, widely modeled as following low-order AR processes, the univariate model is an AR(2), estimated recursively. The univariate model for unemployment is an AR(2) in the change in the unemployment rate, estimated recursively. In the case of inflation, the model is a pseudo-random walk: an AR(4) with no intercept and fixed coefficients of 0.25 on each lag. Point forecasts from this model are as accurate as forecasts from an MA(1) process for the change in inflation, estimated with a rolling window of 40 observations, which Stock and Watson (2007) found to be accurate in point forecasts. The univariate model for the short-term interest rate is an AR(1) in the change in the interest rate, estimated with a rolling sample of 80 observations. Point forecasts from this model are about as accurate as forecasts from a rolling IMA(1) patterned after the Stock and Watson model of inflation.

I report forecasts from conventional constant-variance versions of these AR models and from versions of the models including stochastic volatility. The model of volatility is the same as that described in section 3 for the BVAR, except that the number of model variables is just one in each case. The priors for the volatility components are the same as in the BVAR case. In both the constant variance (AR) and stochastic volatility (AR-SV) cases, forecast distributions are obtained by using MCMC to estimate each model and forecast, with flat priors on the AR coefficients in the models for GDP growth, unemployment, and the interest rate and the AR coefficients fixed at 0.25 in the model for inflation. As with the BVARs, the reported results are based on 10,000 retained draws.

4.2 Simple BVARs

One multivariate forecasting model is a BVAR(4), in GDP growth, the unemployment rate, inflation, and the federal funds rate. The model is estimated with Minnesota priors — specifically the Normal–diffuse prior described in Kadiyala and Karlsson (1997) — via their Gibbs sampling algorithm. The prior means and variances (determined by hyperparameters) are the same as described in section 3.4 for the BVAR-SSPSV model. Flat priors are used for the intercepts of the equations. I consider both recursive and rolling (20 year window) estimates of the model and forecasts. The rolling sample estimation serves as a crude approach to capturing changing shock volatility and allowing gradual change in the VAR coefficients. The number of posterior draws is 15,000, with the first 5000 discarded.

4.3 BVARs with steady state prior (BVAR-SSP)

I also consider forecasts from constant-variance BVAR(4) models with most variables in gap form and an informative prior on the steady state. The model variables consist of GDP growth, the unemployment rate less its trend lagged one period, inflation less the long-run inflation expectation, the funds rate less the long-run inflation expectation, and the change in the long-run inflation expectation. Using Section 3's notation, the model takes the form $\Pi(L)(y_t - \Psi d_t) = v_t$, $v_t \sim N(0, \Sigma)$, with four lags. With a diffuse prior on Σ and the Minnesota and steady state priors described in section 3.4, I estimate the model with the Gibbs sampling approach given in Villani (2009). The estimates and forecasts are obtained from a total of 15,000 draws, with the first 5000 discarded. I consider forecasts from both recursive and rolling (20 year window) estimates of the model.

5 Results

For the models with stochastic volatility to yield density forecasts more accurate than those from models with constant volatilities, it likely needs to be the case that volatility has varied significantly over time. Therefore, as a starting point, it is worth considering the estimates of stochastic volatilities from the BVAR-SSPSV model — specifically, time series of reduced-form residual standard deviations (diagonal elements of $\Sigma_t^{0.5}$) estimated under the recursive scheme. Figure 1 reports estimates (posterior means) obtained with different real-time data vintages. For the key variables of interest, the shaded area provides the volatility time series estimated with data from 1961 through 2008. The lines provide time series estimated with data samples ending in, respectively, 1998:Q4, 1991:Q4, and 1984:Q4 (obtained from data vintages of 1999:Q1, 1992:Q1, and 1985:Q1, respectively). Overall, the estimates confirm significant time variation in volatility, and generally match the contours of estimates shown in such studies as Cogley and Sargent (2005). In particular, volatility fell sharply in the mid-1980s with the Great Moderation. The estimates also reveal a sharp rise in volatility in recent years, reflecting the rise in energy price volatility and the severe recession that started in December 2007.

As might be expected, comparing estimates across real-time data vintages yields some non-trivial differences in volatility estimates. Data revisions — especially benchmark revisions and large annual revisions — lead to some differences across vintages in the stochastic volatility estimates for GDP growth and GDP inflation (a corresponding figure in the not-for-publication appendix using final-vintage data shows much smaller differences across samples). For growth and inflation, the general contours of volatility are very similar across vintages, but levels can differ somewhat. It remains to be seen whether such changes in real time estimates are so great as to make it difficult to improve the calibration of density forecasts by incorporating stochastic volatility. Not surprisingly, with the unemployment and funds rates not revised over time, there are few differences across vintages in the volatility estimates for these variables.

This section proceeds with RMSE results for real-time point forecasts. The following subsections presents results for density forecasts: probabilities of forecasts falling within 70 percent confidence intervals, the tests of Berkowitz (2001) applied to normal transforms of the PITs, and log predictive scores. All of these bear on the calibration of density forecasts (see Mitchell and Vahey (2010) for a recent summary of density calibration). Some additional detail — including mean forecast errors, charts of PITs and normalized forecast errors for a range of models, and illustrative fan charts — is provided in the not-for-publication appendix.

5.1 Point forecasts

Table 1 presents real-time forecast RMSEs for 1985-2008:Q3. The first block of the table reports RMSEs for (constant-variance) AR model forecasts; the remaining blocks report ratios of RMSEs for a given forecast model or method relative to the AR model. In these blocks, entries with value less than 1 mean a forecast is more accurate than the (constant-variance) AR benchmark. To provide a rough measure of statistical significance, Table 2 includes p-values for the null hypothesis that the MSE of a given model is equal to the MSE of the AR benchmark, against the (one-sided) alternative that the MSE of the given model is lower. The not-for-publication appendix provides p-values for tests of equal accuracy of BVAR forecasts against each other (as opposed to against the AR benchmark). The p-values are obtained by comparing Diebold and Mariano (1996)–West (1996) tests against standard normal critical values. Monte Carlo results in Clark and McCracken (2009) indicate that the use of a normal distribution for testing equal accuracy in a finite sample (as opposed to in population, which is the focus of other forecast analyses such as Clark and McCracken (2001)) can be viewed as a conservative guide to inference with models that are nested, as they are here. The standard normal approach tends to be modestly under-sized and have power a little below an asymptotically proper approach, based on a fixed regressor bootstrap that cannot be applied in a BVAR setting.

Consistent with the findings of Clark and McCracken (2008, 2010), the RMSE performance of the conventional BVARs (without variables in gap form and without steady state priors) relative to the benchmark AR models is mixed. For example, at horizons of 1 and 2 quarters and 1 year, the BVAR forecasts often have RMSEs in excess of the AR RMSE. But for growth, unemployment, and the funds rate, the accuracy of BVAR forecasts relative to the univariate forecasts improves as the forecast horizon increases. At the 2-year horizon, BVAR forecasts of these variables are almost always more accurate than AR forecasts, although only for unemployment are the BVAR gains statistically significant. Consider forecasts of unemployment from the recursive BVAR: the RMSE ratio declines from 1.046 at the 1-quarter horizon to 0.989 at the 1-year horizon to 0.722 at the 2-year horizon.

While the pattern is not entirely uniform, for the most part BVARs estimated with rolling samples yield lower RMSEs than BVARs estimated recursively (the pattern is clearer in the set of models with steady state priors). As examples, the RMSE ratios of 1-year ahead forecasts of GDP growth are 1.061 with the recursive BVAR-SSP and 0.958 with the rolling BVAR-SSP, and the RMSE ratios of 1-year ahead forecasts of unemployment are 0.947 and 0.872 with, respectively, the recursive and rolling BVAR-SSP specifications. Admittedly, while the improvements with a rolling scheme are consistent, they are generally too modest to likely be statistically significant.

The BVARs with most variables in gap form and steady state priors (for simplicity, much of the discussion below simply refers to these models as BVARs with steady state priors) generally yield lower RMSEs than conventional BVARs. This finding is in line with evidence in Clark and McCracken (2008, 2010) on the advantage of detrending and evidence in Adolfson, et al. (2007), Beechey and Osterholm (2008), Osterholm (2008), and Wright (2010) on the advantage of steady state priors. The advantage is most striking for 2-year ahead forecasts of inflation. Under a rolling estimation scheme, BVAR and BVAR-SSP forecasts have RMSE ratios of 1.790 and 1.040, respectively. But the advantage, albeit smaller, also applies for most other variables and horizons. At the 1-quarter horizon, rolling BVAR and BVAR-SSP forecasts of GDP growth have RMSE ratios of 1.150 and 1.094, respectively. At the 2-quarter horizon, rolling BVAR and BVAR-SSP forecasts of the funds rate have RMSE ratios of 1.090 and 1.012, respectively. Test p-values provided in the appendix indicate that the forecasts from the rolling BVAR-SSP model are significantly more accurate than the forecasts from the rolling BVAR model, except in the case of unemployment forecasts at all horizons and GDP growth forecasts at the 2-year horizon.

Some of these improvements in RMSEs that come at longer horizons are in part driven by smaller mean errors. As detailed in the appendix, mean errors are often lower (in absolute value) for rolling BVARs than recursively estimated BVARs. Mean errors at longer horizons also tend to be smaller for BVARs with steady state priors than conventional BVARs, especially for inflation and the funds rate.

Adding stochastic volatility to the BVARs with most variables in gap form and steady state priors tends to further improve forecast RMSEs. At the 1-quarter horizon, the recursive BVAR-SSP yields RMSE ratios of 1.160 for GDP growth and 1.144 for the funds rate, while the recursive BVAR-SSPSV yields corresponding ratios of 1.076 and 0.959. At the 1-year horizon, the recursive BVAR-SSP yields RMSE ratios of 1.061 for GDP growth and 0.995 for the funds rate, while the recursive BVAR-SSPSV yields corresponding ratios of 0.983 and 0.914. By the RMSE metric, the rolling BVAR-SSPSV is probably the single best multivariate model. For example, it produces the most instances of rejections of equal accuracy with the AR benchmark. D'Agostino, Gambetti, and Giannone (2009) similarly find that including stochastic volatility in a BVAR (in their case, a model with time-varying parameters) improves the accuracy of point forecasts.

5.2 Density forecasts: interval forecasts

In light of central bank interest in uncertainty surrounding forecasts, confidence intervals, and fan charts, a natural starting point for forecast density evaluation is interval forecasts — that is, coverage rates. Recent studies such as Giordani and Villani (2010) have used interval forecasts as a measure of the calibration of macroeconomic density forecasts. Table 2 reports the frequency with which actual real-time outcomes for growth, unemployment, inflation, and the funds rate fall inside 70 percent highest posterior density intervals estimated in real time with the BVARs (the not-for-publication appendix provides charts of time series of the intervals). Accurate intervals should result in frequencies of about 70 percent. A frequency of more (less) than 70 percent means that, on average over a given sample, the posterior density is too wide (narrow). The table includes p-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on t-statistics. These p-values are provided as a rough gauge of the importance of deviations from correct coverage. The gauge is rough because the theory underlying Christofferson's (1998) test abstracts from forecast model estimation — that is, parameter estimation error — while all forecasts considered in this paper are obtained from estimated models.

As Table 2 shows, the (constant-variance) AR, BVAR, and BVAR-SSP intervals tend to be too wide, with actual outcomes falling inside the intervals much more frequently than the nominal 70 percent rate. For example, for the 1-quarter ahead forecast horizon, the recursive BVAR-SSP coverage rates range from 84.2 to 94.7 percent. Based on the reported p-values, all of these departures from the nominal coverage rate appear to be statistically meaningful. Using the rolling estimation scheme yields slightly to somewhat more accurate interval forecasts (but the departures remain large enough to deliver low p-values, with the exception of the inflation forecasts), with BVAR-SSP coverage rates ranging from 73.7 to 90.5 percent at the 1-step ahead horizon. In some cases, the interval forecasts become more accurate at the 1-year or 2-year horizons, with coverage rates closer to 70 percent. For example, in the case of unemployment forecasts from the rolling BVAR-SSP, the coverage rate improves from 84.2 percent at the 1 quarter horizon to 77.2 at the 1 year horizon.

Adding stochastic volatility to the AR models and to the BVAR with a steady state prior materially improves the calibration of the interval forecasts. For the 1-quarter ahead forecast horizon, the AR-SV coverage rates range from 65.3 to 72.6 percent, down from the AR coverage rate range of 86.3 to 94.7 percent. At the same horizon, the rolling BVAR-SSPSV coverage rates range from 70.5 to 78.9 percent, compared to the rolling BVAR-SSP's range of 73.7 to 90.5 percent. With the BVAR-SSPSV stochastic volatility specifications, for growth, unemployment, and inflation forecasts the p-values for 1-step ahead coverage all exceed 56 percent. But coverage remains too high in the case of the funds rate, at roughly 80 percent — materially better than in the models without stochastic volatility, but still too high. At the 1-year ahead horizon, the rolling BVAR-SSPSV coverage rates range from 70.7 to 79.3 percent, compared to the rolling BVAR-SSP's range of 77.2 to 83.7 percent.

For a given model, differences in coverage across horizons likely reflect a variety of forces, making a single explanation difficult. One force is sampling error: even if a model were correctly specified, random variation in a given data sample could cause the empirical coverage rate to differ from the nominal. Sampling error increases with the forecast horizon, due to the overlap of forecast errors for multi-step horizons (effectively reducing the number of independent observations relative to the one-step horizon). Of course, the increased sampling error across horizons will translate into reduced power to detect departures from accurate coverage.

Another force is the role of (implied or directly estimated) steady states in forecasts at different horizons. As emphasized in such sources as Kozicki and Tinsley (2001a,b), as the horizon increases, forecasts are increasingly determined by the steady states. Some of the apparent improvement in coverage in Table 2 that occurs as the horizon grows (especially for inflation and the funds rate) is due to an increased role of implied or estimated steady states that are too high. Consider, for example, forecasts of the Fed funds rate from the rolling BVAR model. The 1-quarter horizon coverage rate of 92.6 percent indicates the interval forecast is far too wide. However, the model's implied steady state funds rate level is too high. As the horizon increases, the forecasts from the model systematically overstate the funds rate. The bias of the point forecast from the model rises (in absolute value) from -0.160 at the 1-quarter horizon to -1.249 percentage points at the 2-year horizon (not-for-publication appendix). At the 2-year horizon, the forecast interval is likely still too wide, but the whole interval is pushed up by the bias of the point forecasts. As a result, some observations fall below the lower band of the interval, raising the reported coverage rate — but entirely in one tail and not the other. While a total of 29.5 percent of the actual observations fall outside the 70 percent interval at the 2-year horizon, 27.3 percent fall below the lower tail, and only 2.2 percent are above the upper tail. In such cases, of course, the nominal improvement in reported coverage does not actually represent better density calibration. Note that this particular force should not create similar patterns in the log scores, a broader measure of density calibration, reported below.

5.3 Density forecasts: normal transforms of PITs

Normal transforms of PITs can also provide useful indicators of the calibration of density forecasts. The normalized forecast error is defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error and Φ^{-1} is the inverse of the standard normal distribution function. As developed in Berkowitz (2001), the normalized forecast error should be an independent standard normal random variable, because the PIT series should be an independent uniform(0,1) random variable. Berkowitz develops tests based on the normality of the normalized errors that have better power than tests based on the uniformity of the PITs. These tests have been used in recent studies such as Clements (2004) and Jore, Mitchell, and Vahey (2010). Giordani and Villani (2010) also suggest that time series plots of the normalized forecast errors provide useful qualitative evidence of forecast density calibration, and may reveal advantages or disadvantages of a forecast not evident from alternatives such as PIT histograms.

Figure 2 reports time series of normalized forecast errors from the rolling BVAR-SSP and rolling BVAR-SSPSV specifications, with bands representing 90 percent intervals for the normal distribution (normalized errors for other models, consistent with the subset of presented results, are provided in the not-for-publication appendix). Normalized errors from BVARs without stochastic volatility suffer seemingly important departures from the standard normal distribution. Many of the charts indicate the normalized errors have variances well below 1, non-zero means, and serial correlation. The most dramatic examples are for forecasts of the funds rate (from the rolling BVAR-SSP). Less dramatic, although still clear, examples include forecasts of GDP growth and unemployment from the rolling BVAR-SSP. The transforms look best (closest to the standard normal conditions) for forecasts of GDP inflation, which are clearly more variable.

The normalized forecast errors from BVARs with stochastic volatility look much better — with larger variances and means closer to zero. In the case of GDP growth, variability of normalized errors is clearly greater for the BVAR-SSPSV specifications than the BVAR-SSP model, and the mean also looks to be closer to zero. Qualitatively, Giordani and Villani (2010) obtain similar results in comparing forecasts of GDP growth from a constant parameter AR model to forecasts from a model that allows coefficient and variance breaks. However, even with stochastic volatility, there remains an extended period of negative errors in the early 1990s, which implies serial correlation in the normalized errors. The same basic pattern applies to the normalized errors of unemployment forecasts. The results in Figure 2 for inflation forecasts also suggest stochastic volatility improves the behavior of normalized errors, although not as dramatically as with GDP growth and unemployment. Finally, in the case of funds rate forecasts, allowing stochastic volatility also significantly increases the variance of the normalized errors, but seems to leave strong serial correlation.

For a more formal assessment, Table 3 reports various test metrics: the variances of the normalized errors, along with p-values for the null that the variance equals 1; the means of the normalized errors, along with p-values for the null of a zero mean; the AR(1) coefficient estimate and its p-value, obtained by a least squares regression including a constant; and the p-value of Berkowitz's (2001) likelihood ratio test for the joint null of a zero mean, unity variance, and no (AR(1)) serial correlation.

The tests confirm that, without stochastic volatility, variances are materially below 1, means are sometimes non-zero, and serial correlation can be considerable. For example, with the recursive BVAR-SSP model, the variances of the normalized forecast errors range from 0.205 (funds rate) to 0.636 (inflation), with p-values close to 0. With the same model, the AR(1) coefficients are 0.310 for GDP growth, 0.393 for unemployment, -0.198 for inflation, and 0.674 for the funds rate; the corresponding p-values are all close to zero, except in the case of inflation, for which the p-value is 0.051. However, particularly in terms of means and variances, the rolling scheme fares somewhat better than the recursive. Not surprisingly, given results such as these for means, variances, and AR(1) coefficients, the p-values of the Berkowitz (2001) test are nearly zero for constant variance AR, recursive and rolling BVAR, and recursive and rolling BVAR-SSP forecasts, with the exception of rolling forecasts of GDP inflation.

By the formal metrics, as by the charts, allowing stochastic volatility improves the calibration of density forecasts. In the case of the recursive BVAR-SSPSV specification, the variances of the normalized forecast errors range from 0.848 (federal funds rate) to 1.030 (inflation), with p-values of 0.223 or more. The AR(1) coefficients are all lower (in absolute value) for forecasts from the recursive BVAR-SSPSV than from the recursive BVAR-SSP specification. For unemployment and inflation, the p-values of the Berkowitz (2001) test are above 10 percent. For GDP growth, the p-value of the test exceeds 5 percent. Adding stochastic volatility to AR models yields a qualitatively similar improvement (relative to constant-variance AR models) in the properties of normalized forecast errors, with the forecasts passing the Berkowitz test for all but the funds rate, for which the violation appears to be due to serial correlation in the normalized error.

5.4 Density forecasts: log predictive density scores

The overall calibration of the density forecasts can most broadly measured with log predictive density scores, used in such recent studies as Geweke and Amisano (2010). For computational tractability, I compute the log predictive density score based on the Gaussian (quadratic) formula given in Adolfson, Linde, and Villani (2005), under which a lower score implies a better model. For brevity, I report only average log scores for the full vector of variables of interest (GDP growth, unemployment, inflation, and the funds rate). The not-for-publication appendix provides scores for each individual variable.

To help provide a rough gauge of the significance of score differences, I rely on the methodology developed in Amisano and Giacomini (2007), and report p-values for selected differences in mean scores, under the null of a zero mean. Because the theoretical basis for the test provided by Amisano and Giacomini requires forecasts estimated with rolling samples of data (but does take account of the effects of uncertainty associated with the estimation of model parameters), I only apply the test to pairs of forecasts from models estimated with the rolling scheme: BVAR against BVAR-SSP, BVAR against BVAR-SSPSV, and BVAR-SSP against BVAR-SSPSV.

The average log predictive density scores reported in Table 4 show that a rolling estimation scheme almost always yields better (lower) log scores than does a recursive estimation scheme, although sometimes by small amounts. For example, in the case of the BVAR-SSP model, the rolling scheme yields log scores of 8.526 at the 1 quarter horizon and 10.372 at the 1 year horizon, while the recursive scheme yields corresponding log scores of 8.772 and 10.732. In addition, using gap forms for most variables and a steady state prior usually improves log scores: given the estimation scheme (recursive or rolling), log scores are typically lower for the BVAR-SSP model than the BVAR. Continuing with the same example, the rolling BVAR has log scores of 8.600 at the 1 quarter horizon and 10.767 at the 1 year horizon (compared to the rolling BVAR-SSP's scores of 8.526 and 10.327, respectively).

Allowing stochastic volatility offers improvements in log scores that seem especially considerable at short horizons. The log scores of the rolling BVAR-SSPSV are 7.345 (1Q), 9.716 (2Q), 10.019 (1Y), and 12.380 (2Y), compared to the rolling BVAR-SSP model's log scores of 8.526, 10.519, 10.372, and 13.021, respectively. Including stochastic volatility in AR models also significantly improves log scores relative to constant-variance AR models. For instance, at the 1-quarter horizon, the scores of the AR and AR-SV models are 9.228 and 7.596, respectively. For the vector of four variables being forecast, the BVARs with stochastic volatility score better than the AR models with stochastic volatility, especially at longer horizons. This difference is likely due to covariances among forecasts that the BVARs better capture. The appendix indicates that, for each individual variable separately, the

AR-SV and BVAR-SSPSV scores are broadly comparable.

The p-values of the differences in average log scores reported in the lower panel of Table 4 indicate that the improvements in density forecasts from BVARs associated with stochastic volatility are statistically meaningful at short horizons, although mixed at longer horizons. In comparing the (rolling in all cases) BVAR-SSP against the BVAR-SSPSV, the differences in average log scores are: 1.180, with a p-value of 0.000 (1Q); 0.803, with a p-value of 0.021 (2Q); 0.352, with a p-value of 0.393 (1Y); and 0.641, with a p-value of 0.106 (2Y). On the basis of this evidence, it seems reasonable to conclude that modeling stochastic volatility significantly improves the calibration of density forecasts, although more convincingly at shorter horizons than longer horizons.

6 Conclusions

Central banks and other forecasters have become increasingly interested in various aspects of density forecasts. However, sharp changes in macroeconomic volatility — such as the Great Moderation and the more recent rise in volatility associated with greater variation in energy prices and the deep recession — pose significant challenges to density forecasting. Accordingly, this paper examines, with real-time data, density forecasts of GDP growth, unemployment, inflation, and the federal funds rate from BVAR models with stochastic volatility. Drawing on past research, to improve the accuracy of point forecasts, the model of interest includes most variables in a gap, or deviation from trend, form and incorporates informative priors on the steady states of the model variables.

The evidence presented in the paper shows that adding stochastic volatility to the BVAR with most variables in gap form and a steady state prior materially improves the real-time accuracy of density forecasts and more modestly improves the accuracy of point forecasts. The paper also shows that adding stochastic volatility to univariate AR models materially improves density forecast calibration relative to AR models with constant variances. The density evidence includes interval forecasts (coverage rates), time series plots and various tests applied to normal transforms of the probability integral transforms, and log predictive density scores. In the absence of stochastic volatility, models estimated with rolling samples of data are more accurate in density forecasting than models estimated recursively. But modeling stochastic volatility yields larger gains in forecast accuracy.

References

Adolfson, Malin, Michael K. Andersson, Jesper Linde, Mattias Villani, and Anders Vredin (2007), "Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks," *International Journal of Central Banking*, December, 111-144.

Adolfson, Malin, Jesper Linde, and Mattias Villani (2005), "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," Sveriges Riksbank Working Paper No. 190.

Amisano, Gianni, and Raffaella Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics* 25, 177-190.

Beechey, Meredith, and Par Osterholm (2008), "A Bayesian Vector Autoregressive Model with Informative Steady-State Priors for the Australian Economy," *Economic Record* 84, 449-465.

Berkowitz, Jeremy (2001), "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business and Economic Statistics* 19, 465-474.

Chatfield, Chris, Anne B. Koehler, J. Keith Ord, and Ralph D. Snyder (2001), "A New Look at Models for Exponential Smoothing," *The Statistician* 50 (Part 2), 147-159.

Christoffersen, Peter F. (1998), "Evaluating Interval Forecasts," *International Economic Review* 39, 841-862.

Clark, Todd E. (2009), "Is the Great Moderation Over? An Empirical Analysis," *Federal Reserve Bank of Kansas City Economic Review* 94 (no.4), 5-42.

Clark, Todd E. and Michael W. McCracken (2010), "Averaging Forecasts from VARs with Uncertain Instabilities," *Journal of Applied Econometrics* 25, 5-29.

Clark, Todd E., and Michael W. McCracken (2009), "Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy," Working Paper 2009-05A, Federal Reserve Bank of St. Louis, October 2009.

Clark, Todd E. and Michael W. McCracken (2008), "Forecasting with Small Macroeconomic VARs in the Presence of Instability," in *Forecasting in the Presence of Structural Breaks and Model Uncertainty*, Rapach DE, Wohar ME (eds). Emerald Publishing.

Clark, Todd E., and Michael W. McCracken (2001), "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics* 105, 85-110.

Clements, Michael P. (2004), "Evaluating the Bank of England Density Forecasts of Inflation," *Economic Journal* 114, 844-866.

Cogley, Timothy (2002), 'A Simple Adaptive Measure of Core Inflation," *Journal of Money, Credit, and Banking* 34, 94-113.

Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005), "Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System," *Journal of Economic Dynamics and Control* 29, 1893-1925.

Cogley Timothy, and Thomas J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post-World War II U.S.," *Review of Economic Dynamics* 8, 262-302.

Croushore, Dean (2006), "Forecasting with Real-Time Macroeconomic Data," *Handbook of Economic Forecasting*, G. Elliott, C. Granger, and A. Timmermann (Eds.), North Holland: Amsterdam.

Croushore, Dean and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-30.

D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2009), "Macroeconomic Forecasting and Structural Change," ECARES Working Paper 2009-020.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Del Negro, Marco, and Frank Schorfheide (2004), "Priors from General Equilibrium Models for VARs," *International Economic Review* 45, 643- xxx

Faust, Jon, and Jonathan H. Wright (2009), "Comparing Greenbook and Reduced Form Forecasts using a Large Realtime Dataset," *Journal of Business and Economic Statistics* 27, 468-479.

Geweke, John, and Gianni Amisano (2010), "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, forthcoming.

Giordani, Paolo, and Mattias Villani (2010), "Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction," *International Journal of Forecasting*, forthcoming.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics* 12, 371418.

Kadiyala, K. Rao and Sune Karlsson (1997), "Numerical Methods for Estimation and Inference in Bayesian VAR Models," *Journal of Applied Econometrics* 12, 99-132.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies* 65, 361-393.

Kozicki, Sharon and Peter A. Tinsley (2001a), "Shifting Endpoints in the Term Structure of Interest Rates," *Journal of Monetary Economics* 47, 613-652.

Kozicki, Sharon and Peter A. Tinsley (2001b), "Term Structure Views of Monetary Policy under Alternative Models of Agent Expectations," *Journal of Economic Dynamics and Control* 25, 149-84.

Jore, Anne Sofie, James Mitchell, and Shaun P. Vahey (2010), "Combining Forecast Densities from VARS with Uncertain Instabilities," *Journal of Applied Econometrics*, forthcoming.

Makridakis, Spyros, and Michele Hibon (2000), "The M3-Competition: Results, Conclusions, and Implications," *International Journal of Forecasting* 16, 451-476.

Mitchell, James, and Kenneth F. Wallis (2010), "Evaluating Density Forecasts: Forecast Combinations, Model Mixtures, Calibration, and Sharpness," *Journal of Applied Econometrics*, forthcoming.

Osterholm, Par (2008), "Can Forecasting Performance Be Improved by Considering the Steady State? An Application to Swedish Inflation and Interest Rate," *Journal of Forecasting* 27, 41-51.

Primiceri, Giorgio (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72, 821-852

Romer, Christina D., and David H. Romer (2000), "Federal Reserve Information and the Behavior of Interest Rates," *American Economic Review* 90, 429-457.

Sims, Christopher A. (2002), "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity* 2, 1-40.

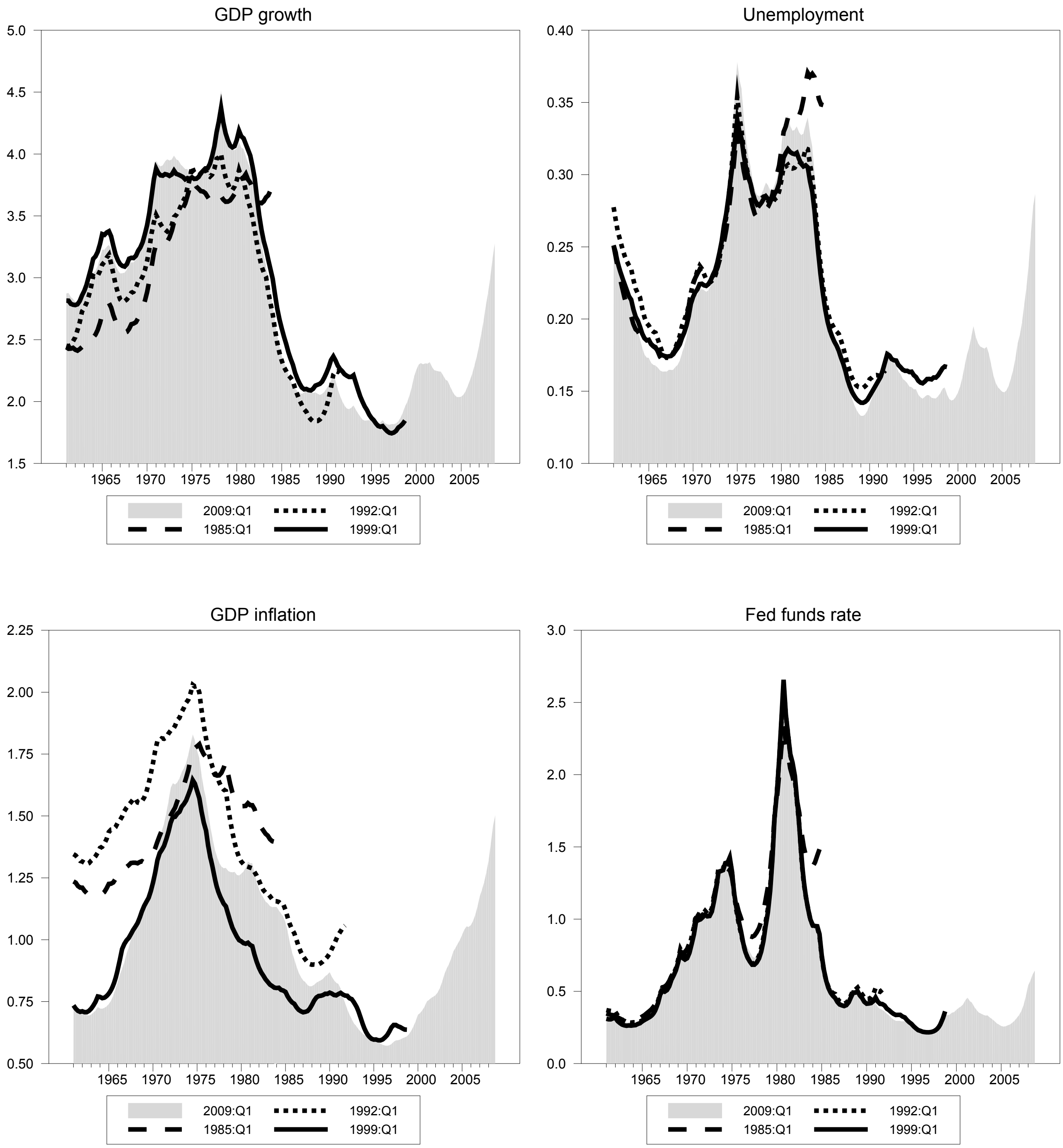
Stock, James H., and Mark W. Watson (2007), "Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit, and Banking* 39, 3-33.

Villani, Mattias (2009), "Steady-State Priors for Vector Autoregressions," *Journal of Applied Econometrics* 24, 630-650.

West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084.

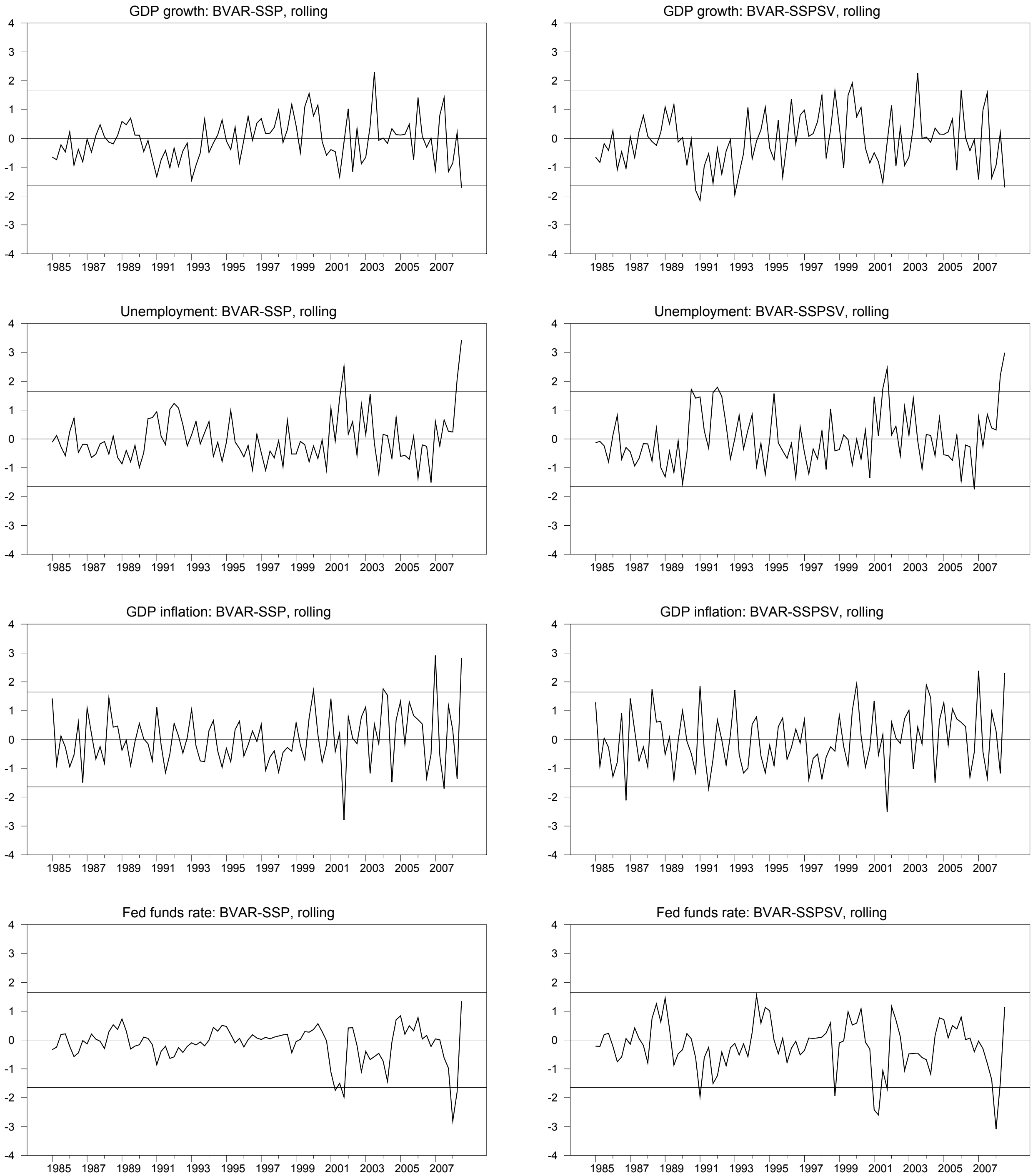
Wright, Jonathan H. (2010), "Evaluating Real-Time VAR Forecasts with an Informative democratic Prior," manuscript, Johns Hopkins University.

Figure 1. Posterior Means of Residual Standard Deviations, Real-Time Vintages
BVAR-SSPSV, recursive



Notes: The figure reports posterior means of the time series of estimates of the reduced-form residual variances in the BVAR-SSPSV model (recursive), estimated at various points in time with the vintage of data indicated. The dates given for each line (1985:Q1, 1992:Q1, etc.) correspond to the dates of the data vintages; at each of these points in time, the model was estimated with the indicated vintage, using data through the prior quarter.

Figure 2. Normalized Forecast Errors from Selected Models
horizon = 1Q



Notes: The normalized forecast errors shown are defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the probability integral transform of a one-step ahead forecast error (generated in real time) and Φ^{-1} is the inverse of the standard normal distribution function. The horizontal lines included in the charts represent 90 percent intervals for the normal distribution.

Table 1. Real-Time Forecast RMSEs, 1985-2008Q3

(RMSEs for benchmark AR models in first panel, RMSE ratios in all others)

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
		AR		
GDP growth	1.783	1.832	1.320	1.445
Unemployment	0.174	0.313	0.572	1.055
GDP inflation	0.965	1.001	0.669	0.932
Fed funds rate	0.416	0.808	1.489	2.463
		AR-SV		
GDP growth	1.023 (0.998)	1.023 (0.990)	1.047 (0.988)	1.073 (0.971)
Unemployment	0.996 (0.383)	1.005 (0.588)	1.016 (0.630)	1.026 (0.629)
GDP inflation	1.002 (0.819)	1.001 (0.585)	0.999 (0.343)	0.999 (0.275)
Fed funds rate	0.896 (0.000)	0.943 (0.041)	1.019 (0.700)	1.086 (0.944)
		BVAR, recursive		
GDP growth	1.242 (0.997)	1.218 (0.986)	1.215 (0.900)	0.924 (0.279)
Unemployment	1.046 (0.728)	1.031 (0.637)	0.991 (0.471)	0.722 (0.002)
GDP inflation	1.102 (0.982)	1.141 (0.993)	1.307 (0.998)	1.573 (0.998)
Fed funds rate	1.157 (0.993)	1.130 (0.920)	1.047 (0.680)	0.964 (0.395)
		BVAR, rolling		
GDP growth	1.150 (0.989)	1.133 (0.947)	1.086 (0.715)	0.908 (0.318)
Unemployment	1.052 (0.738)	1.015 (0.553)	0.934 (0.295)	0.663 (0.001)
GDP inflation	1.116 (0.981)	1.206 (0.995)	1.461 (0.997)	1.790 (0.994)
Fed funds rate	1.145 (0.997)	1.090 (0.882)	1.021 (0.595)	0.971 (0.405)
		BVAR-SSP, recursive		
GDP growth	1.160 (0.983)	1.127 (0.942)	1.061 (0.670)	0.899 (0.230)
Unemployment	1.045 (0.729)	1.011 (0.555)	0.947 (0.316)	0.719 (0.013)
GDP inflation	1.051 (0.864)	1.018 (0.634)	1.044 (0.674)	1.078 (0.692)
Fed funds rate	1.144 (0.984)	1.112 (0.881)	0.995 (0.483)	0.859 (0.137)
		BVAR-SSP, rolling		
GDP growth	1.094 (0.950)	1.049 (0.759)	0.958 (0.382)	0.849 (0.175)
Unemployment	1.008 (0.550)	0.957 (0.283)	0.872 (0.109)	0.674 (0.013)
GDP inflation	1.044 (0.825)	1.017 (0.626)	1.014 (0.561)	1.040 (0.641)
Fed funds rate	1.079 (0.944)	1.012 (0.579)	0.900 (0.071)	0.787 (0.013)
		BVAR-SSPSV, recursive		
GDP growth	1.076 (0.896)	1.058 (0.797)	0.983 (0.449)	0.884 (0.229)
Unemployment	0.981 (0.374)	0.941 (0.158)	0.901 (0.138)	0.735 (0.012)
GDP inflation	1.034 (0.787)	1.018 (0.650)	0.995 (0.475)	0.994 (0.478)
Fed funds rate	0.959 (0.169)	0.953 (0.217)	0.914 (0.143)	0.829 (0.061)
		BVAR-SSPSV, rolling		
GDP growth	1.059 (0.887)	1.030 (0.692)	0.940 (0.311)	0.863 (0.191)
Unemployment	0.981 (0.377)	0.927 (0.119)	0.855 (0.042)	0.688 (0.003)
GDP inflation	1.037 (0.791)	1.029 (0.717)	1.003 (0.513)	0.981 (0.412)
Fed funds rate	0.936 (0.044)	0.914 (0.035)	0.863 (0.013)	0.782 (0.009)

Notes:

1. In each quarter t from 1985:Q1 through 2008:Q3, vintage t data (which end in $t-1$) are used to form forecasts for periods t ($h = 1Q$), $t+1$ ($h = 2Q$), $t+3$ ($h = 1Y$), and $t+7$ ($h = 2Y$). The forecasts of GDP growth and inflation for the $h = 1Y$ and $h = 2Y$ horizons correspond to annual percent changes: average growth and average inflation from t through $t+3$ and $t+4$ through $t+7$, respectively. The forecast errors are calculated using the second–available (real–time) estimates of growth and inflation as the actual data, and currently available measures of unemployment and the federal funds rate as actuals.
2. The entries in the first panel are RMSEs, for variables defined in annualized percentage points. All other entries are RMSE ratios, for the indicated model relative to the corresponding (constant-variance) AR model.
3. p-values of t-tests of equal MSE, taking the AR models with constant volatilities as the benchmark, are given in parentheses. These are one-sided Diebold-Mariano-West tests, of the null of equal forecast accuracy against the alternative that the non-benchmark model in question is more accurate. The standard errors entering the test statistics are computed with the Newey-West estimator, with a bandwidth of 0 at the 1-quarter horizon and 1.5 × horizon in the other cases.

Table 2. Real-Time Forecast Coverage Rates, 1985-2008Q3

(Frequencies of actual outcomes falling inside 70% intervals)

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
	AR			
GDP growth	0.947 (0.000)	0.936 (0.000)	0.913 (0.000)	0.909 (0.000)
Unemployment	0.884 (0.000)	0.936 (0.000)	0.935 (0.000)	0.898 (0.000)
GDP inflation	0.863 (0.000)	0.862 (0.000)	0.880 (0.000)	0.955 (0.000)
Fed funds rate	0.947 (0.000)	0.915 (0.000)	0.891 (0.000)	0.864 (0.013)
	AR-SV			
GDP growth	0.716 (0.733)	0.734 (0.549)	0.717 (0.820)	0.716 (0.861)
Unemployment	0.684 (0.741)	0.766 (0.157)	0.772 (0.293)	0.739 (0.641)
GDP inflation	0.653 (0.332)	0.670 (0.555)	0.696 (0.930)	0.670 (0.674)
Fed funds rate	0.726 (0.565)	0.691 (0.880)	0.554 (0.067)	0.625 (0.340)
	BVAR, recursive			
GDP growth	0.863 (0.000)	0.883 (0.000)	0.674 (0.755)	0.852 (0.001)
Unemployment	0.874 (0.000)	0.851 (0.005)	0.739 (0.677)	0.682 (0.854)
GDP inflation	0.811 (0.006)	0.830 (0.001)	0.870 (0.000)	0.795 (0.132)
Fed funds rate	0.947 (0.000)	0.915 (0.000)	0.837 (0.043)	0.784 (0.359)
	BVAR, rolling			
GDP growth	0.821 (0.002)	0.851 (0.000)	0.707 (0.934)	0.852 (0.022)
Unemployment	0.821 (0.002)	0.819 (0.020)	0.717 (0.837)	0.636 (0.533)
GDP inflation	0.758 (0.188)	0.755 (0.198)	0.783 (0.089)	0.659 (0.550)
Fed funds rate	0.926 (0.000)	0.883 (0.000)	0.793 (0.205)	0.705 (0.965)
	BVAR-SSP, recursive			
GDP growth	0.884 (0.000)	0.894 (0.000)	0.804 (0.122)	0.909 (0.000)
Unemployment	0.863 (0.000)	0.840 (0.020)	0.783 (0.245)	0.795 (0.128)
GDP inflation	0.842 (0.000)	0.894 (0.000)	0.913 (0.000)	0.943 (0.000)
Fed funds rate	0.947 (0.000)	0.904 (0.000)	0.826 (0.050)	0.739 (0.720)
	BVAR-SSP, rolling			
GDP growth	0.853 (0.000)	0.894 (0.000)	0.815 (0.052)	0.920 (0.000)
Unemployment	0.842 (0.000)	0.830 (0.006)	0.772 (0.213)	0.795 (0.243)
GDP inflation	0.737 (0.415)	0.809 (0.010)	0.826 (0.028)	0.841 (0.122)
Fed funds rate	0.905 (0.000)	0.904 (0.000)	0.837 (0.023)	0.739 (0.689)
	BVAR-SSPSV, recursive			
GDP growth	0.705 (0.910)	0.745 (0.387)	0.696 (0.957)	0.739 (0.644)
Unemployment	0.705 (0.910)	0.702 (0.969)	0.674 (0.726)	0.591 (0.250)
GDP inflation	0.695 (0.911)	0.745 (0.389)	0.848 (0.003)	0.773 (0.305)
Fed funds rate	0.800 (0.015)	0.670 (0.630)	0.576 (0.110)	0.568 (0.240)
	BVAR-SSPSV, rolling			
GDP growth	0.705 (0.910)	0.755 (0.268)	0.707 (0.923)	0.727 (0.802)
Unemployment	0.726 (0.565)	0.723 (0.667)	0.717 (0.813)	0.705 (0.955)
GDP inflation	0.716 (0.733)	0.723 (0.579)	0.793 (0.074)	0.795 (0.195)
Fed funds rate	0.789 (0.032)	0.734 (0.589)	0.728 (0.706)	0.625 (0.479)

Notes:
1. See the notes to Table 1.
2. The table reports the frequencies with which actual outcomes fall within 70 percent bands computed from the posterior distribution of forecasts.
3. The table includes in parentheses p-values for the null of correct coverage (empirical = nominal rate of 70 percent), based on t-statistics using standard errors computed with the Newey-West estimator, with a bandwidth of 0 at the 1-quarter horizon and 1.5 $\times$ horizon in the other cases.

Table 3. Tests of Normalized Errors of 1-Step Ahead Real-Time Forecasts, 1985-2008Q3

	variance (p-value)	mean (p-value)	AR(1) coef. (p-value)	LR test p-value
	AR			
GDP growth	0.269 (0.000)	-0.089 (0.133)	0.060 (0.537)	0.000
Unemployment	0.433 (0.000)	0.016 (0.823)	-0.150 (0.235)	0.000
GDP inflation	0.482 (0.000)	0.007 (0.911)	0.071 (0.516)	0.000
Fed funds rate	0.278 (0.000)	-0.025 (0.738)	0.523 (0.002)	0.000
	AR-SV			
GDP growth	0.832 (0.139)	-0.164 (0.111)	0.052 (0.611)	0.179
Unemployment	0.960 (0.772)	0.125 (0.254)	-0.133 (0.226)	0.356
GDP inflation	1.069 (0.596)	-0.001 (0.990)	0.119 (0.252)	0.677
Fed funds rate	0.974 (0.888)	-0.094 (0.476)	0.333 (0.003)	0.009
	BVAR, recursive			
GDP growth	0.506 (0.000)	-0.334 (0.001)	0.337 (0.000)	0.000
Unemployment	0.527 (0.000)	0.162 (0.151)	0.401 (0.000)	0.000
GDP inflation	0.672 (0.001)	-0.169 (0.004)	-0.204 (0.054)	0.002
Fed funds rate	0.200 (0.000)	-0.165 (0.022)	0.667 (0.000)	0.000
	BVAR, rolling			
GDP growth	0.534 (0.000)	-0.107 (0.320)	0.266 (0.008)	0.000
Unemployment	0.675 (0.070)	0.140 (0.254)	0.381 (0.007)	0.000
GDP inflation	0.901 (0.512)	-0.103 (0.228)	-0.143 (0.172)	0.332
Fed funds rate	0.400 (0.000)	-0.192 (0.026)	0.537 (0.002)	0.000
	BVAR-SSP, recursive			
GDP growth	0.443 (0.000)	-0.226 (0.021)	0.310 (0.001)	0.000
Unemployment	0.555 (0.001)	0.096 (0.406)	0.393 (0.001)	0.000
GDP inflation	0.636 (0.000)	-0.065 (0.269)	-0.198 (0.051)	0.006
Fed funds rate	0.205 (0.000)	-0.159 (0.034)	0.674 (0.000)	0.000
	BVAR-SSP, rolling			
GDP growth	0.517 (0.000)	-0.051 (0.600)	0.204 (0.047)	0.000
Unemployment	0.648 (0.037)	-0.008 (0.942)	0.332 (0.034)	0.001
GDP inflation	0.880 (0.453)	0.027 (0.726)	-0.163 (0.097)	0.369
Fed funds rate	0.407 (0.000)	-0.136 (0.136)	0.546 (0.001)	0.000
	BVAR-SSPSV, recursive			
GDP growth	0.852 (0.223)	-0.158 (0.215)	0.189 (0.065)	0.063
Unemployment	0.921 (0.614)	0.024 (0.856)	0.203 (0.079)	0.269
GDP inflation	1.030 (0.805)	-0.089 (0.265)	-0.124 (0.206)	0.540
Fed funds rate	0.848 (0.361)	-0.225 (0.097)	0.494 (0.000)	0.000
	BVAR-SSPSV, rolling			
GDP growth	0.851 (0.186)	-0.081 (0.509)	0.182 (0.063)	0.172
Unemployment	0.878 (0.446)	0.015 (0.910)	0.266 (0.026)	0.075
GDP inflation	1.003 (0.980)	-0.014 (0.868)	-0.123 (0.184)	0.710
Fed funds rate	0.759 (0.201)	-0.188 (0.116)	0.483 (0.000)	0.000

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

Table 4. Real-Time Forecast Average Log Scores, 1985-2008Q3

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
	Average scores for forecasts of all four variables			
AR	9.228	11.726	12.513	15.410
AR-SV	7.596	10.600	12.180	15.417
BVAR, recursive	8.860	11.005	11.092	13.734
BVAR, rolling	8.600	10.731	10.767	13.582
BVAR-SSP, recursive	8.772	10.811	10.732	13.286
BVAR-SSP, rolling	8.526	10.519	10.372	13.021
BVAR-SSPSV, recursive	7.343	9.957	10.515	12.959
BVAR-SSPSV, rolling	7.345	9.716	10.019	12.380
	Differences in mean scores of rolling BVARs (p-values)			
BVAR vs. BVAR-SSP	0.075 (0.122)	0.212 (0.018)	0.395 (0.036)	0.561 (0.044)
BVAR vs. BVAR-SSPSV	1.255 (0.000)	1.016 (0.006)	0.748 (0.102)	1.202 (0.027)
BVAR-SSP vs. BVAR-SSPSV	1.180 (0.000)	0.803 (0.021)	0.352 (0.393)	0.641 (0.106)

Notes:
1. See the notes to Table 1.
2. The entries in the upper panel of the table are average values of log predictive density scores, computed with the Gaussian (quadratic) formula given in Adolfson, Linde, and Villani (2005), under which a lower score implies a better model.
3. The entries in the lower panel of the table are differences in average log predictive density scores and p-values from Amisano and Giacomini (2007) tests of equal average scores. The tests and p-values are computed by regressions of differences in log scores (time series) on a constant, using the Newey-West estimator of the variance of the regression constant (with a bandwidth of 0 at the 1-quarter horizon and 1.5×horizon in the other cases). All of the test results are based on forecasts from models estimated with rolling samples of data, per the assumptions of Amisano and Giacomini (2007).

Not-for-Publication Appendix to "Real-Time Density Forecasts from BVARs with Stochastic Volatility" *

Todd E. Clark
Federal Reserve Bank of Kansas City

March 2010

*Economic Research Dept.; Federal Reserve Bank of Kansas City; 1 Memorial Drive; Kansas City, MO 64198; `todd.e.clark@kc.frb.org`; phone: (816)881-2575; fax: (816)881-2199.

1 Overview

This not-for-publication appendix provides additional detail on data used in the paper and some additional results. The additional results consist of: charts of volatility estimates based on just the final vintage of data; a table of mean forecast errors; a table of Diebold-Mariano-West tests of equal accuracy of point forecasts from the BVARs; charts of PITs and normalized forecast errors for a wider range of models than the paper's Chart 2 includes; charts of time series of 70 percent interval forecasts; detailed results on log predictive scores; and illustrative fan charts.

2 Data

This section details the construction of the long-run inflation expectations series used in the BVAR estimation and forecasting.

For each quarterly vintage, I construct a real-time long-run expectations series by combining exponential smoothing-based estimates for 1960:Q1-1979:Q3 with Blue Chip expectations data for 1979:Q4 through the forecast origin.

The Blue Chip data are the 6-10 year ahead (in some surveys, 5-9 year ahead) expectations of inflation in the GDP (or GNP) deflator or price index. For most of the sample, the results were published in the March and October surveys (on the 10th of each month). But for a few earlier months, the results were published in May or November. To ensure the data are available in real time, the March figures are treated as being available in the Q2 vintage, but not the Q1 vintage. Similarly, the October figures are treated as being available in the Q4 vintage, but not the Q3 vintage.

The exponentially smoothed series are constructed in real time, for each quarterly vintage t, as follows. (1) Initialize the filter with the mean rate of price increase for 1953-59, using that vintage's price series. If data are not available for that 1953-59 period, use the same mean from the most recent vintage with data available for that period. The mean becomes the exponentially smoothed estimate for period 1959:4. Use exponential smoothing with a smoothing coefficient of .05, to estimate trend inflation from 1960:1 through t-1. Define the value of long-run expected inflation for period t as the exponentially smoothed trend estimated with data through t-1. Because the first vintage of price data is 1965:Q4, this approach begins with that vintage. To make things as real time as possible, estimates of trend inflation for 1960:Q31-1965:Q3 are also those obtained by exponential smoothing of the 1965:Q4 vintage of data. Note, however, that applying exponential smoothing to the historical data of the 2008:Q4 vintage yields a very similar trend estimate.

For each vintage t starting in 1985:Q1, a time series on inflation expectations is compiled as follows. For 1960:Q1 through 1979:Q3, the expectation is estimated with the real time exponentially smoothed series. For 1979:Q4 through t, the expectation is based on Blue Chip. For odd-numbered quarters before period t (in vintage t), I linearly interpolate between quarters 2 and 4. When period t of vintage t is odd-numbered, the expectation is simply the Blue Chip value from period t-1.

3 Volatility estimates

Appendix Figure 1 reports final-vintage estimates of stochastic volatilities from the BVAR-SSPSV model — specifically, time series of residual standard deviations (posterior means of the time series of standard deviations, the diagonal elements of $\Sigma_t^{0.5}$) estimated under the recursive scheme. These estimates are based on just the last available vintage of data, the 2009:Q1 vintage. For each variable (except that, for brevity, the chart focuses on the primary variables of interest and omits the volatility estimates for the change in the long-run inflation expectation), the shaded area provides the volatility time series estimated with data from 1961 through 2008. The lines provide time series estimated with data samples ending in, respectively, 1998:Q4, 1991:Q4, and 1984:Q4 (using pseudo-vintages dated 1999:Q1, 1992:Q1, and 1985:Q1).

Across estimates generated with different data samples, the volatilities are generally very similar. Not surprisingly, the biggest revisions occur with volatilities estimated at the end of a sample (the biggest differences in lines occur at the end of the lines) — only being able to do one-sided filtering at the end of a sample has some modest effect. For example, in the case of GDP growth, for volatility in 1990, the estimate obtained with a data sample ending in 1991 (blue line) exceeds the estimate obtained with a data sample ending in 2008.

As indicated in the paper, volatility estimates based on real-time data for GDP growth and inflation yield larger differences across vintages, due to data revisions, particularly those associated with annual and benchmark revisions.

4 Mean forecast errors

Appendix Table 1 presents real-time mean forecast errors for 1985-2008:Q3. The statistical significance of the mean forecast errors can be gauged by the p-values included in the table, obtained by simple regressions of the errors on a constant.

The mean errors of forecasts for growth, inflation, and the funds rate are consistently negative, indicating the forecasts are consistently too high. On this dimension, the rolling VARs often fare better than the recursively estimated VARs, with smaller (in absolute value) mean errors. For example, in the case of 1-year ahead forecasts of GDP growth, the mean error from the recursive BVAR-SSP is -0.545, while the mean error from the rolling BVAR is -0.063. In general, the average errors in forecasts of unemployment are materially smaller than the average errors for other variables, but are more mixed in sign (sometimes positive, sometimes negative). For instance, with the recursive BVAR-SSP model, the 1-year ahead forecast errors average -0.545 for GDP growth and 0.055 for unemployment.

At longer horizons, the BVARs with most variables in gap form and informative steady state priors tend to fare better than the conventional BVARs, especially for inflation and the funds rate. For 2-year ahead projections of inflation, the rolling BVAR yields an average error of -1.093, compared to the rolling BVAR-SSP's mean error of -0.432. The BVARs with steady state priors and stochastic volatility tend to yield average errors broadly comparable to the BVARs with just steady state priors — generally a bit lower in the recursive case but sometimes a bit higher in the rolling case. In the case of 1-quarter ahead forecasts of GDP growth, the recursive (rolling) BVAR-SSP has an average error of -0.705 (-0.190), while the recursive (rolling) BVAR-SSPSV has an average error of -0.389 (-0.237). Finally, in many but not all cases, the mean errors are smaller (in absolute value) for the AR forecasts than the VAR forecasts.

5 PITs results

The probability integral transform (PIT) emphasized by Diebold, Gunther, and Tay (1998) provides a more general indicator of the accuracy of density intervals. For each variable and forecast horizons of 1 quarter and 1 year, Appendix Figures 2-9 present PIT histograms, obtained as decile counts of PIT transforms. For optimal density forecasts at the 1-step horizon, the PIT series would be independent uniform (0,1) random variables. Accordingly, the histograms would be flat (with 9.5 observations per bin at the 1-quarter horizon).

Studies such as Christoffersen and Mazzotta (2005), Clements (2004), and Geweke and Amisano (2010) consider similar measures of density forecasts. To provide some measure of a gauge of the importance of departures from the iid uniform distribution, the 1-quarter ahead PITs charts include 90 percent intervals estimated under the binomial distribution (following Diebold, Gunther, and Tay (1998)). These intervals are only intended as a rough guide; among other issues, the intervals abstract from the possible effects of model parameter estimation on the large-sample distributions of PITs.

Forecasts from BVARs without stochastic volatility suffer seemingly large departures from uniformity. While some might see the PITs for rolling BVARs as looking a bit flatter than those from recursive BVARs, the differences are pretty small, with both suffering material departures from uniformity. In the case of 1-quarter ahead GDP growth and unemployment rate forecasts, the PITs have too much mass in the middle of the distribution. The PITs for inflation are somewhat flatter, with more modest crossings of the 90 percent bands. The departures from uniformity are most severe for the PITs of funds rate forecasts; the histograms of the BVAR-SSP forecast PITs look more like normal densities than uniform densities. The clustering of mass in the middle of the distributions of the PITs most likely reflects estimated forecast distributions that are too wide, because the forecast models without stochastic volatility treat the residual variances as constant and therefore, following the Great Moderation, over-estimate volatilities.

Forecasts from BVARs with stochastic volatility look to be much closer to being uniformly distributed. For the recursive and rolling BVAR-SSPSV forecasts, the PITs are quite a bit flatter than for the models without stochastic volatility; the BVAR-SSPSV PITs are much less prone to crossing the 90 percent bands. The BVAR-SSPSV PITs exceed the threshold twice for the Fed funds rate, once for GDP growth, and never for unemployment and inflation.

6 Density forecasts: normal transforms of PITs

Normal transforms of PITs can also provide useful indicators of the accuracy of density forecasts. The normalized forecast error is defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error and Φ^{-1} is the inverse of the standard normal distribution function. As developed in Berkowitz (2001), the normalized forecast error should be an independent standard normal random variable, because the PIT series should be an independent uniform(0,1) random variable. Giordani and Villani (2010) suggest that time series plots of the normalized forecast errors provide useful qualitative evidence of forecast density accuracy, and may reveal advantages or disadvantages of a forecast not evident from alternatives such as PIT histograms.

Appendix Figures 10-13 report time series of the normalized forecast errors, with bands representing 90 percent intervals for the normal distribution. In line with the PITs results, normalized errors from BVARs without stochastic volatility suffer seemingly important departures from the standard normal distribution. Many of the charts indicate the normalized errors have variances well below 1, non-zero means, and serial correlation. The most dramatic examples are for forecasts of the funds rate — such as from the recursive BVAR-SSP. Less dramatic, although still clear, examples include forecasts of GDP growth and unemployment from the recursive BVAR-SSP. The transforms look best (closest to the standard normal conditions) for forecasts of GDP inflation, which are clearly more variable. In some cases, results seem to look a bit better — at least in the sense of having a larger variance — under the rolling estimation scheme than the recursive, but qualitatively similar.

The normalized forecast errors from BVARs with stochastic volatility look much better — with larger variances and means closer to zero. In the case of GDP growth, variability of normalized errors is clearly greater for the BVAR-SSPSV specifications than the BVAR-SSP or BVAR models, and the mean also looks to be closer to zero. However, even with stochastic volatility, there remains an extended period of negative errors in the early 1990s, which implies serial correlation in the errors. The same basic pattern applies to the normalized errors of unemployment forecasts. The results in Appendix Figure 12 for inflation forecasts also suggest stochastic volatility improves the behavior of normalized errors, although not as dramatically as with GDP growth and unemployment. Finally, in the case of funds rate forecasts, allowing stochastic volatility also significantly increases the variance of the normalized errors, but seems to leave strong serial correlation.

7 Interval forecasts

Appendix Figures 14-17 provide time series of 70% interval forecasts, obtained from the recursive BVAR-SSP (pair of black lines), rolling BVAR-SSP (pair of blue lines), and rolling BVAR-SSPSV (pair of green lines) models.[1]

[1] These intervals based on percentiles represent equal-tail credible sets.

Consistent with the coverage statistics provided in the paper (Table 2), the charts show that the forecast intervals are generally quite different across methods (plotting intervals for the other models considered in the paper show similarly considerable differences). At shorter horizons, the differences across methods appear greater for GDP growth and the fed funds rate than for the other variables. In the case of unemployment, the differences in interval width at each moment in time are actually greater than the reported chart suggests, because the scale of the chart is large enough (driven by what amount to changes in trend inflation) relative to uncertainty at each moment in time to obscure differences in the intervals at each point in time. When the *width* of the intervals is plotted, the results for unemployment generally mirror those for other variables. But at the longer horizons (most notably at the 2-year horizon) covered in Appendix Figures 16-17, the differences across methods are more visibly considerable for unemployment and inflation. This pattern likely reflects the greater importance of the steady state means in longer-run forecasts than in shorter-run forecasts.

The differences in methods also vary over time. Until roughly the late 1990s, the interval forecasts from the recursive and rolling BVAR-SSP specifications tend to be fairly similar (with some exceptions at the 2-year forecast horizon), but quite different from the interval forecast based on the rolling BVAR-SSPSV specification. In most cases, the intervals from the model with stochastic volatility are considerably narrower than the intervals from the models without stochastic volatility. For roughly the last decade of the sample, the interval forecasts from the rolling BVAR-SSP and rolling BVAR-SSPSV models tend to be fairly similar, while the interval forecasts from the recursive BVAR-SSP specification tend to be considerably wider than the intervals from the other two methods. These broad patterns reflect the ability of each method to pick up the large changes in volatility highlighted in the paper's Figure 1 and Appendix Figure 1. For example, by the late 1980s, the model with stochastic volatility detects the reduction in volatility associated with the Great Moderation and yields relatively narrow intervals, while the models without stochastic volatility produce wider intervals based on (assumed constant-variance) samples that include a significant period of high volatility. As another example, by late 2007, the period of high volatility has passed out of the sample used to estimate the rolling BVAR-SSP model, while it remains in the sample used to estimate the recursive BVAR-SSP specification.

8 Fan chart illustration

To further illustrate the practical consequences of time variation in conditional volatilities, Appendix Figures 18 and 19 provide fan charts of forecasts made in the middle of, respectively, 1995:Q4 and 2008:Q3, using the 1995:Q4 and 2008:Q3 vintages of data from the RTDSM, with data samples ending in 1995:Q3 and 2008:Q2. Following Cogley, Morozov, and Sargent (2005), the figures report percentiles of the marginal density for each variable at each horizon.[2] In keeping with common practice, the GDP growth and inflation forecasts are reported as forecasts of four-quarter averages. Consequently, for these variables, the probability bands widen as the horizon increases from the current quarter (for which three quarters of growth and inflation entering the four-quarter average are known), to the next quarter (for which two quarters of growth and inflation entering the four-quarter average are known), and so on. In the interest of brevity, results are reported for only two of the models or methods: the BVAR-SSP and BVAR-SSPSV specifications estimated with rolling samples of data.

In general, allowing for stochastic volatility significantly affects the fan chart estimates. As noted in the paper, volatility has risen sharply in recent quarters (including 2008:Q4 and 2009:Q1, not included in the estimates underlying these fan charts). The simple rolling sample estimates that treat error variances as constant in the 20-year sample can only very gradually capture such changes. The model with stochastic volatility can more rapidly pick up the changes in error variances. As a consequence, in the 1995:Q4 example, the fan charts are much narrower for the BVAR-SSPSV specification than the BVAR-SSP specification, indicating much less uncertainty. In the 2008:Q3 example, there remain differences across the specification, but the differences are more modest than in the 1995:Q4 example. In this later example, for unemployment, inflation, and the federal funds rate, the estimated fan chart bands are wider — conveying more uncertainty surrounding the outlook — for the model with stochastic volatility (BVAR-SSPSV, rolling) than the model without (BVAR-SSP, rolling). In the case of GDP growth forecasts, though, the fan charts are quite similar for the two estimates.

[2]These fan charts based on percentiles represent equal-tail credible sets.

References

Amisano, Gianni, and Raffaella Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests," *Journal of Business and Economic Statistics* 25, 177-190.

Adolfson, Malin, Jesper Linde, and Mattias Villani (2005), "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," Sveriges Riksbank Working Paper No. 190.

Berkowitz, Jeremy (2001), "Testing Density Forecasts, With Applications to Risk Management," *Journal of Business and Economic Statistics* 19, 465-474.

Christoffersen, Peter, and Stefano Mazzotta (2005), "The Accuracy of Density Forecasts from Foreign Exchange Options," *Journal of Financial Econometrics* 3, 578-605.

Clements, Michael P. (2004), "Evaluating the Bank of England Density Forecasts of Inflation," *Economic Journal* 114, 844-866.

Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005), "Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System," *Journal of Economic Dynamics and Control* 29, 1893-1925.

Diebold, Francis X., Todd A. Gunther, and Anthony S. Tay (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review* 39, 863-883.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.

Geweke, John, and Gianni Amisano (2010), "Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns," *International Journal of Forecasting*, forthcoming.

Giacomini, Rafaella, and Halbert White (2006), "Tests of Conditional Predictive Ability," *Econometrica* 74, 1545-1578.

Giordani, Paolo, and Mattias Villani (2010), "Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction," *International Journal of Forecasting*, forthcoming.

West, Kenneth D. (1996), "Asymptotic Inference About Predictive Ability," *Econometrica* 64, 1067-1084.

Appendix Table 1. Real-Time Mean Forecast Errors, 1985-2008Q3

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
		AR		
GDP growth	-0.313 (0.082)	-0.347 (0.133)	-0.332 (0.182)	-0.405 (0.180)
Unemployment	0.003 (0.881)	0.001 (0.974)	-0.022 (0.847)	-0.093 (0.727)
GDP inflation	0.006 (0.949)	-0.014 (0.900)	-0.008 (0.946)	-0.021 (0.912)
Fed funds rate	-0.036 (0.396)	-0.086 (0.511)	-0.161 (0.597)	-0.157 (0.784)
		AR-SV		
GDP growth	-0.363 (0.048)	-0.415 (0.080)	-0.403 (0.119)	-0.512 (0.125)
Unemployment	0.023 (0.195)	0.053 (0.219)	0.111 (0.335)	0.225 (0.398)
GDP inflation	0.004 (0.971)	-0.016 (0.888)	-0.009 (0.937)	-0.023 (0.905)
Fed funds rate	-0.056 (0.136)	-0.150 (0.189)	-0.351 (0.226)	-0.638 (0.270)
		BVAR, recursive		
GDP growth	-1.040 (0.000)	-1.076 (0.000)	-0.930 (0.002)	-0.562 (0.045)
Unemployment	0.039 (0.033)	0.084 (0.113)	0.160 (0.211)	0.159 (0.488)
GDP inflation	-0.227 (0.034)	-0.497 (0.000)	-0.556 (0.000)	-1.203 (0.000)
Fed funds rate	-0.178 (0.000)	-0.369 (0.011)	-0.693 (0.021)	-1.221 (0.016)
		BVAR, rolling		
GDP growth	-0.355 (0.086)	-0.323 (0.296)	-0.213 (0.504)	0.126 (0.690)
Unemployment	0.032 (0.082)	0.061 (0.246)	0.091 (0.448)	0.009 (0.963)
GDP inflation	-0.197 (0.070)	-0.432 (0.002)	-0.499 (0.004)	-1.093 (0.006)
Fed funds rate	-0.160 (0.001)	-0.338 (0.013)	-0.668 (0.019)	-1.249 (0.014)
		BVAR-SSP, recursive		
GDP growth	-0.705 (0.000)	-0.678 (0.017)	-0.545 (0.065)	-0.242 (0.441)
Unemployment	0.020 (0.276)	0.040 (0.453)	0.055 (0.658)	-0.034 (0.883)
GDP inflation	-0.090 (0.388)	-0.280 (0.003)	-0.320 (0.000)	-0.745 (0.000)
Fed funds rate	-0.164 (0.000)	-0.341 (0.018)	-0.634 (0.029)	-1.073 (0.024)
		BVAR-SSP, rolling		
GDP growth	-0.190 (0.339)	-0.122 (0.651)	-0.063 (0.819)	0.052 (0.860)
Unemployment	-0.008 (0.669)	-0.021 (0.662)	-0.067 (0.535)	-0.209 (0.288)
GDP inflation	-0.009 (0.930)	-0.125 (0.238)	-0.156 (0.181)	-0.432 (0.105)
Fed funds rate	-0.084 (0.065)	-0.182 (0.164)	-0.372 (0.165)	-0.686 (0.129)
		BVAR-SSPSV, recursive		
GDP growth	-0.389 (0.043)	-0.388 (0.146)	-0.300 (0.286)	-0.127 (0.692)
Unemployment	0.006 (0.722)	0.012 (0.804)	0.010 (0.929)	-0.071 (0.761)
GDP inflation	-0.069 (0.503)	-0.217 (0.024)	-0.241 (0.011)	-0.557 (0.001)
Fed funds rate	-0.110 (0.005)	-0.250 (0.037)	-0.503 (0.058)	-0.871 (0.064)
		BVAR-SSPSV, rolling		
GDP growth	-0.237 (0.217)	-0.194 (0.448)	-0.129 (0.621)	0.009 (0.977)
Unemployment	0.004 (0.813)	0.004 (0.937)	-0.013 (0.903)	-0.103 (0.619)
GDP inflation	-0.011 (0.913)	-0.107 (0.308)	-0.125 (0.271)	-0.335 (0.164)
Fed funds rate	-0.085 (0.029)	-0.201 (0.071)	-0.415 (0.087)	-0.726 (0.094)

Notes:
1. See the notes to Table 1 in the paper.
2. The entries are mean forecast errors (actuals less forecasts), for variables defined in annualized percentage points. The entries in parentheses are p-values for the null of zero mean error, based on t-statistics using Newey-West standard errors with bandwidth of 0 at horizon 1 and bandwidth of 1.5*horizon in other cases.

Appendix Table 2. Diebold-Mariano-West Tests of Equal MSEs of BVAR Forecasts, Real-Time Forecasts, 1985-2008Q3

(t-tests, with p-values)

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
	BVAR vs. BVAR-SSP, both rolling			
GDP growth	1.964 (0.025)	2.464 (0.007)	2.696 (0.004)	1.240 (0.107)
Unemployment	1.224 (0.110)	0.835 (0.202)	0.645 (0.260)	-0.158 (0.563)
GDP inflation	2.265 (0.012)	3.636 (0.000)	3.284 (0.001)	2.886 (0.002)
Fed funds rate	2.282 (0.011)	1.792 (0.037)	2.142 (0.016)	2.355 (0.009)
	BVAR vs. BVAR-SSPSV, both rolling			
GDP growth	2.632 (0.004)	2.913 (0.002)	2.608 (0.005)	0.872 (0.192)
Unemployment	1.876 (0.030)	1.276 (0.101)	0.912 (0.181)	-0.486 (0.686)
GDP inflation	2.341 (0.010)	3.447 (0.000)	3.207 (0.001)	2.775 (0.003)
Fed funds rate	3.809 (0.000)	2.810 (0.002)	2.520 (0.006)	2.426 (0.008)
	BVAR-SSP vs. BVAR-SSPSV, both rolling			
GDP growth	2.539 (0.006)	1.103 (0.135)	0.526 (0.299)	-0.971 (0.834)
Unemployment	1.708 (0.044)	1.216 (0.112)	0.512 (0.304)	-0.328 (0.628)
GDP inflation	0.722 (0.235)	-0.894 (0.814)	0.545 (0.293)	1.284 (0.100)
Fed funds rate	3.283 (0.001)	2.621 (0.004)	2.175 (0.015)	0.503 (0.307)

Notes:
1. The table entries are Diebold-Mariano-West t-tests (p-values) of equal MSEs, taking the first model listed for each panel as the benchmark and the second as the alternative. The variances entering the test statistics use the Newey-West estimator, with a bandwidth of 0 at the 1-quarter horizon and 1.5×horizon in the other cases. All of the results in the table are based on forecasts from models estimated with rolling samples of data, as required in the theoretical results of Giacomini and White (2006) that provide an asymptotic basis for comparing the tests against the standard normal distribution.
2. Positive test statistics indicate that the alternative model (the second in each listed pair) has a lower MSE than the null model (the first in each listed pair). Because the models are nested, the reported p-values are for one-sided tests of the null of equal accuracy against the alternative that the second model in each listed pair is more accurate.

Appendix Table 3. Real-Time Forecast Average Log Scores, 1985-2008Q3

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
	All four variables			
AR	9.228	11.726	12.513	15.410
AR-SV	7.596	10.600	12.180	15.417
BVAR, recursive	8.860	11.005	11.092	13.734
BVAR, rolling	8.600	10.731	10.767	13.582
BVAR-SSP, recursive	8.772	10.811	10.732	13.286
BVAR-SSP, rolling	8.526	10.519	10.372	13.021
BVAR-SSPSV, recursive	7.343	9.957	10.515	12.959
BVAR-SSPSV, rolling	7.345	9.716	10.019	12.380
	GDP growth			
AR	4.613	4.677	3.879	4.007
AR-SV	4.152	4.247	3.656	3.911
BVAR, recursive	4.628	4.648	3.812	3.732
BVAR, rolling	4.470	4.496	3.520	3.632
BVAR-SSP, recursive	4.567	4.584	3.602	3.771
BVAR-SSP, rolling	4.433	4.444	3.392	3.599
BVAR-SSPSV, recursive	4.245	4.334	3.514	3.517
BVAR-SSPSV, rolling	4.205	4.276	3.413	3.485
	Unemployment			
AR	-0.342	0.945	2.174	3.277
AR-SV	-0.525	0.739	2.051	3.403
BVAR, recursive	-0.360	0.675	1.733	2.319
BVAR, rolling	-0.344	0.629	1.602	2.201
BVAR-SSP, recursive	-0.371	0.651	1.673	2.408
BVAR-SSP, rolling	-0.391	0.581	1.537	2.308
BVAR-SSPSV, recursive	-0.605	0.504	1.600	2.371
BVAR-SSPSV, rolling	-0.572	0.525	1.548	2.307
	Inflation			
AR	2.945	3.018	2.320	3.049
AR-SV	2.787	2.908	2.209	2.888
BVAR, recursive	3.014	3.272	2.834	3.739
BVAR, rolling	3.008	3.222	2.767	3.636
BVAR-SSP, recursive	2.944	3.129	2.599	3.350
BVAR-SSP, rolling	2.920	3.053	2.451	3.145
BVAR-SSPSV, recursive	2.810	2.923	2.257	3.001
BVAR-SSPSV, rolling	2.813	2.926	2.258	2.933
	Federal funds rate			
AR	2.013	3.088	4.145	5.078
AR-SV	1.179	2.716	4.262	5.230
BVAR, recursive	2.224	3.107	3.935	4.662
BVAR, rolling	2.058	3.052	3.970	4.753
BVAR-SSP, recursive	2.189	3.050	3.817	4.424
BVAR-SSP, rolling	2.025	2.989	3.811	4.407
BVAR-SSPSV, recursive	1.090	2.528	3.952	4.634
BVAR-SSPSV, rolling	1.140	2.440	3.646	4.432

Notes:
1. See the notes to Table 1 in the paper.
2. The table entries are average values of log predictive density scores, computed with the Gaussian (quadratic) formula given in Adolfson, Linde, and Villani (2005), under which a lower score implies a better model.

Appendix Table 4. Amisano-Giacomini Test Applied to Average Log Scores of Real-Time Forecasts, 1985-2008Q3

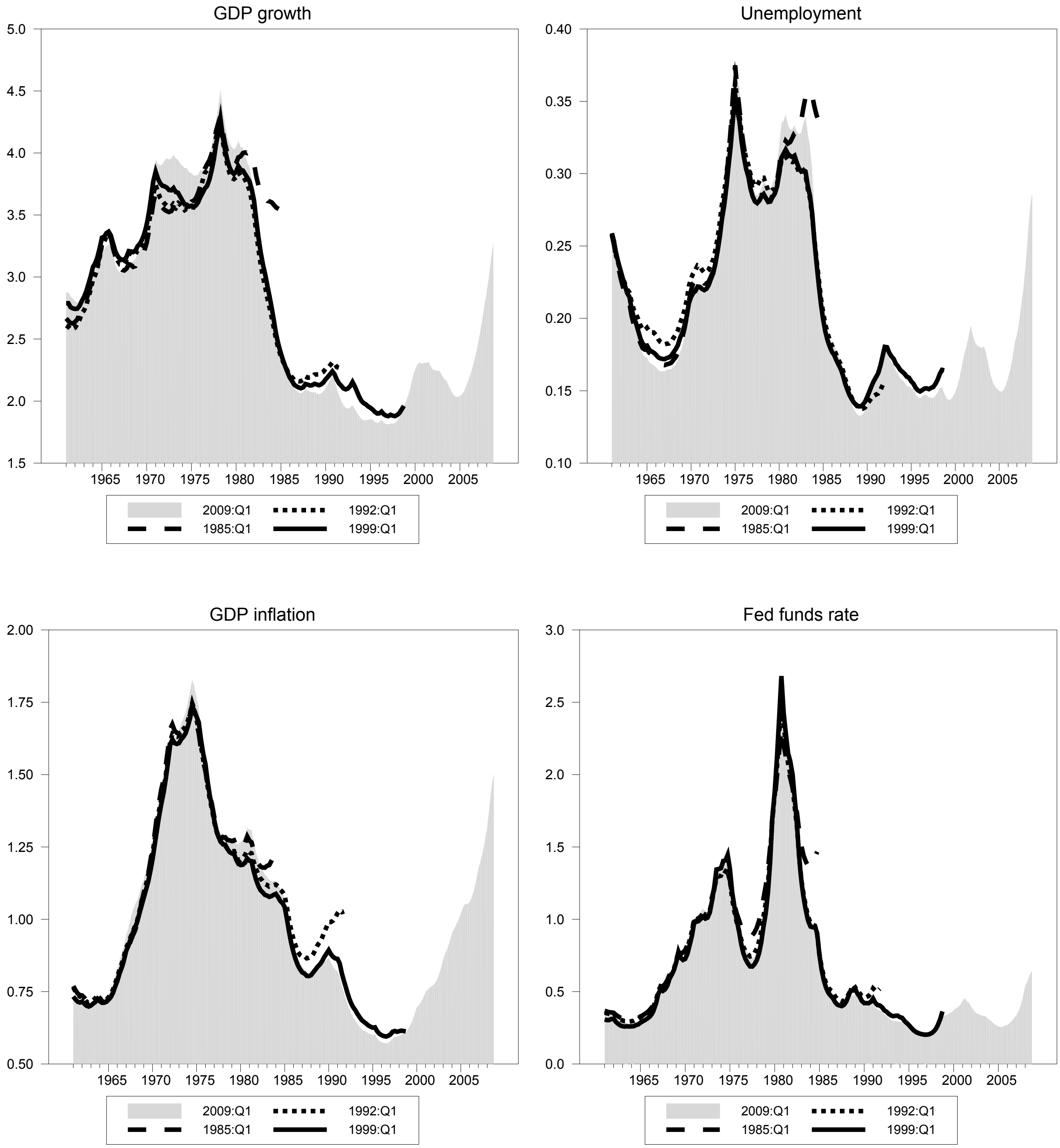
(Mean Differences in Log Scores, with p-values)

	$h = 1Q$	$h = 2Q$	$h = 1Y$	$h = 2Y$
	All four variables			
BVAR vs. BVAR-SSP	0.075 (0.122)	0.212 (0.018)	0.395 (0.036)	0.561 (0.044)
BVAR vs. BVAR-SSPSV	1.255 (0.000)	1.016 (0.006)	0.748 (0.102)	1.202 (0.027)
BVAR-SSP vs. BVAR-SSPSV	1.180 (0.000)	0.803 (0.021)	0.352 (0.393)	0.641 (0.106)
	GDP growth			
BVAR vs. BVAR-SSP	0.037 (0.054)	0.052 (0.019)	0.128 (0.048)	0.033 (0.537)
BVAR vs. BVAR-SSPSV	0.266 (0.000)	0.220 (0.010)	0.107 (0.309)	0.147 (0.095)
BVAR-SSP vs. BVAR-SSPSV	0.229 (0.000)	0.168 (0.068)	-0.021 (0.867)	0.114 (0.325)
	Unemployment			
BVAR vs. BVAR-SSP	0.047 (0.082)	0.049 (0.502)	0.065 (0.709)	-0.108 (0.503)
BVAR vs. BVAR-SSPSV	0.228 (0.000)	0.104 (0.242)	0.053 (0.715)	-0.106 (0.418)
BVAR-SSP vs. BVAR-SSPSV	0.181 (0.001)	0.055 (0.578)	-0.012 (0.933)	0.002 (0.989)
	Inflation			
BVAR vs. BVAR-SSP	0.089 (0.009)	0.169 (0.000)	0.315 (0.000)	0.492 (0.001)
BVAR vs. BVAR-SSPSV	0.196 (0.001)	0.296 (0.000)	0.509 (0.000)	0.703 (0.005)
BVAR-SSP vs. BVAR-SSPSV	0.107 (0.037)	0.127 (0.012)	0.194 (0.024)	0.211 (0.071)
	Federal funds rate			
BVAR vs. BVAR-SSP	0.033 (0.029)	0.063 (0.006)	0.159 (0.000)	0.346 (0.000)
BVAR vs. BVAR-SSPSV	0.919 (0.000)	0.613 (0.003)	0.325 (0.172)	0.321 (0.188)
BVAR-SSP vs. BVAR-SSPSV	0.886 (0.000)	0.550 (0.007)	0.165 (0.510)	-0.025 (0.928)

Notes:

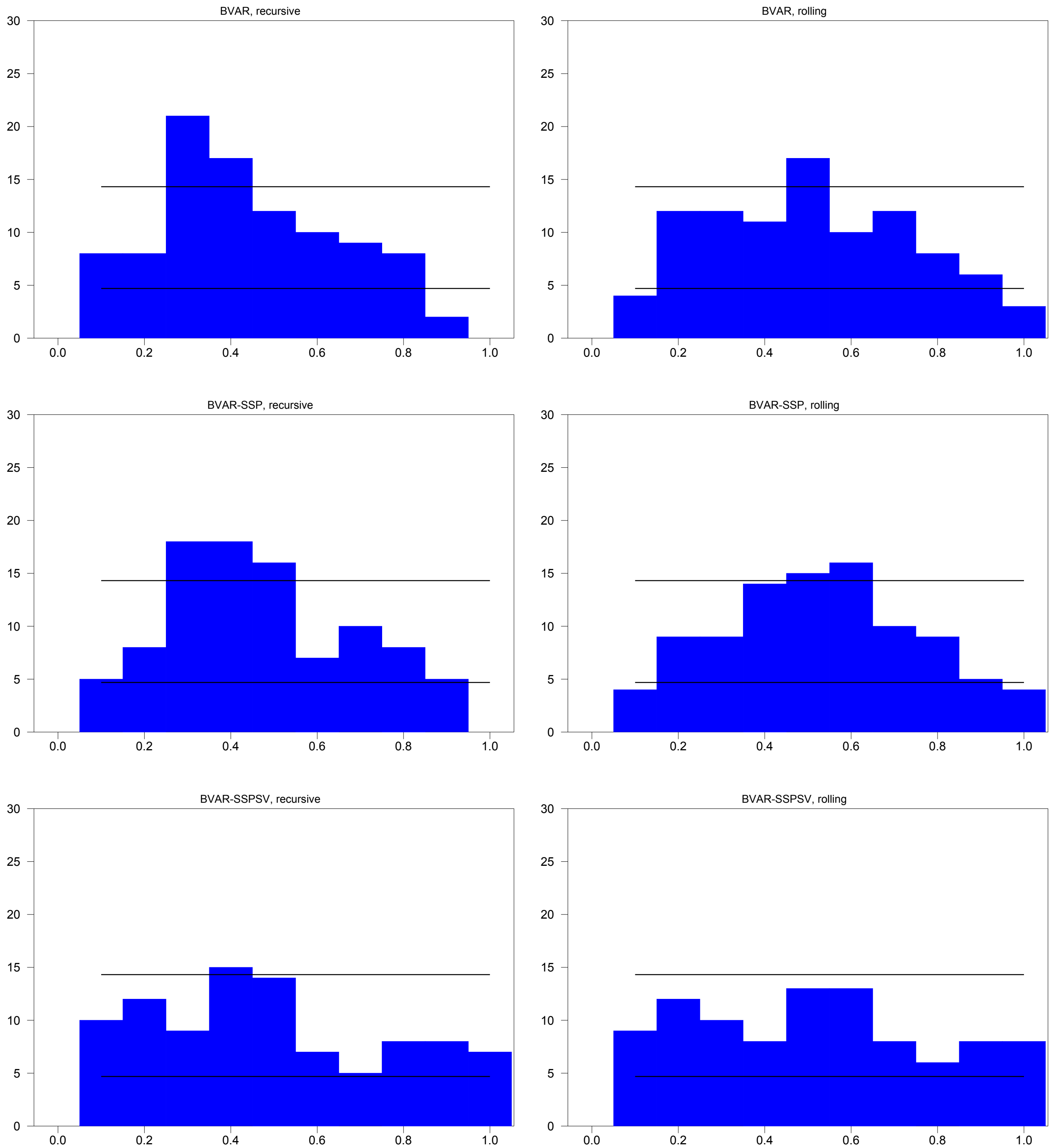
1. The table entries are differences in average log predictive density scores and p-values from Amisano and Giacomini (2007) tests of equal average scores. The tests and p-values are computed by regressions of differences in log scores (time series) on a constant, using the Newey-West estimator of the variance of the regression constant (with a bandwidth of 0 at the 1-quarter horizon and 1.5×horizon in the other cases). All of the results in the table are based on forecasts from models estimated with rolling samples of data.

Appendix Figure 1. Posterior Means of Residual Standard Deviations, Final Vintage Data
BVAR-SSPSV, recursive



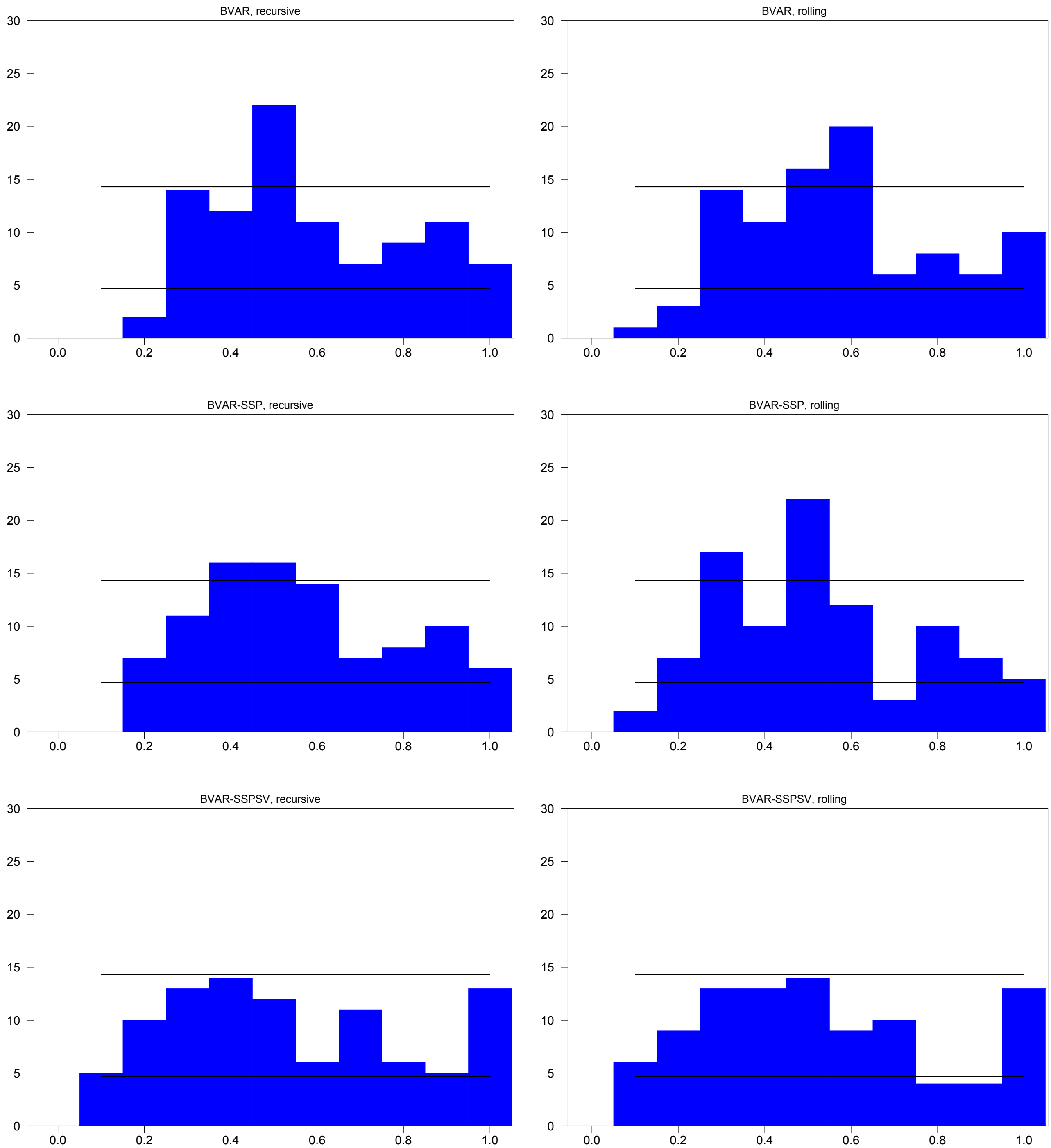
Notes: The figure reports posterior means of the time series of estimates of the reduced-form residual variances in the BVAR-SSPSV model (recursive), estimated at various points in time with the 2009:Q1 vintage of data, which includes data through 2008:Q4. The dates given for each line (1985:Q1, 1992:Q1, etc.) correspond to pseudo-vintage dates; at each of these points in time, the model was estimated with data through the prior quarter.

Appendix Figure 2. PIT histogram for GDP growth
horizon = 1Q



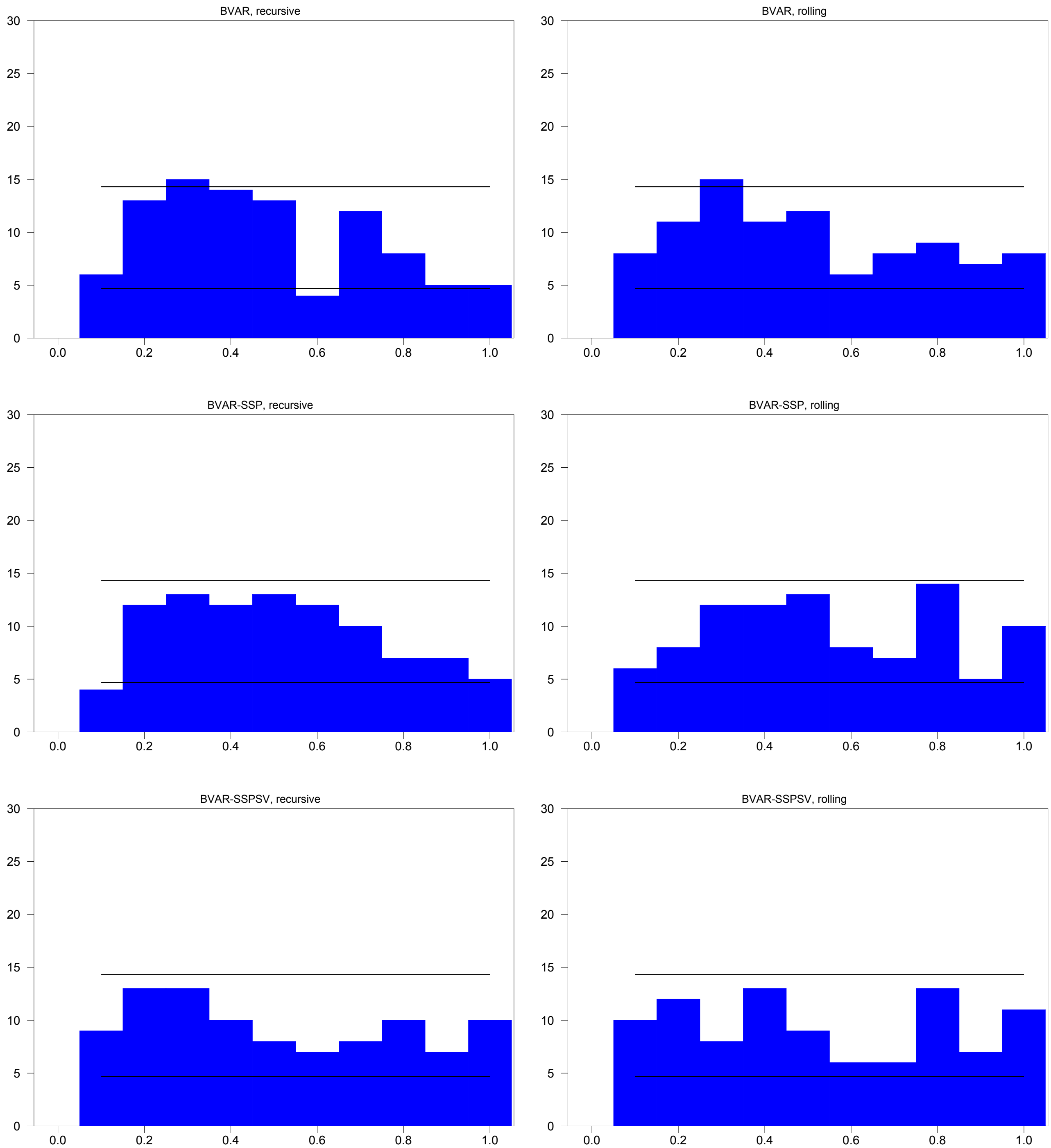
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts, with 90 percent intervals (intended only as a rough guide) estimated under the binomial distribution.

Appendix Figure 3. PIT histogram for Unemployment
horizon = 1Q



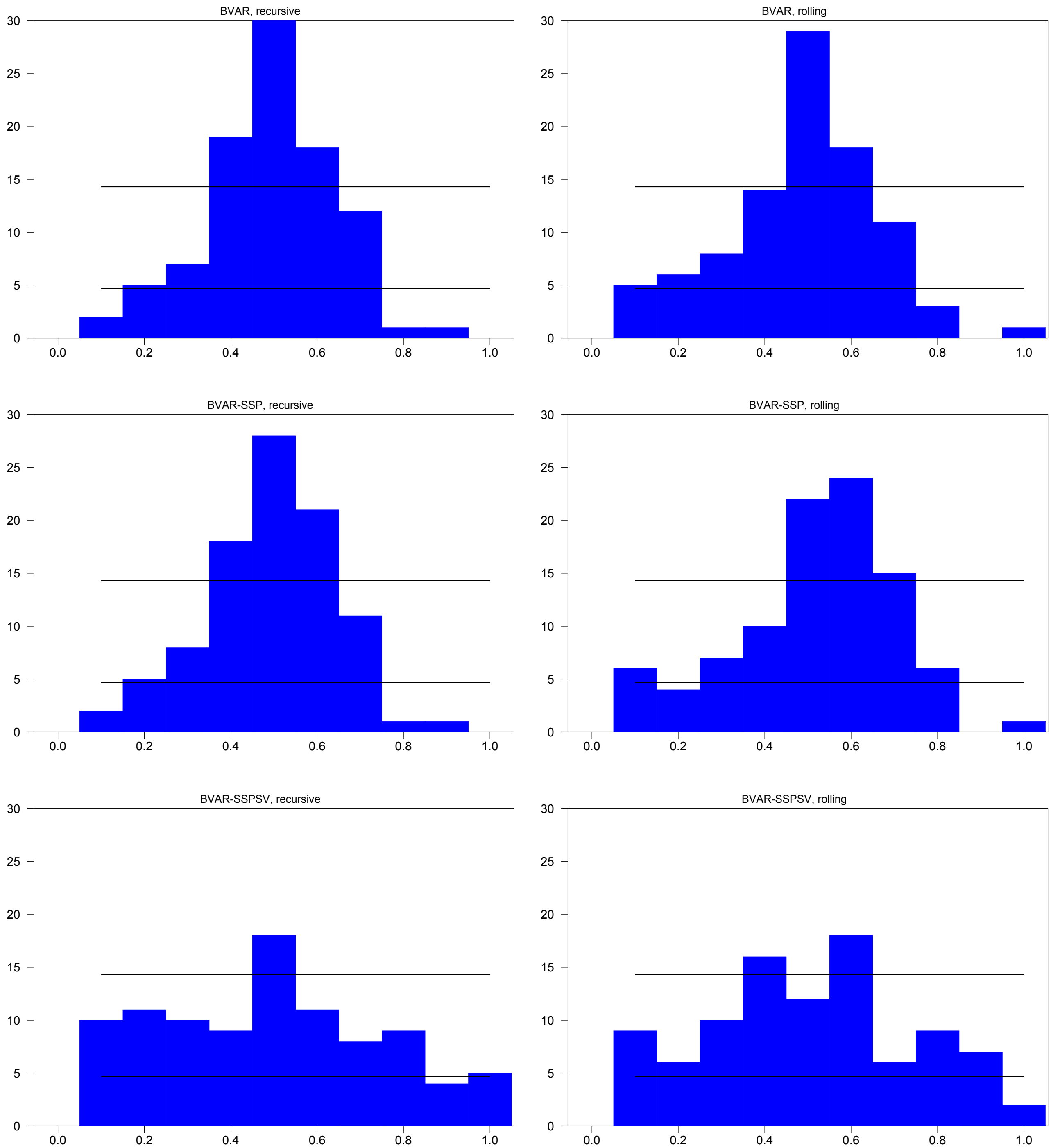
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts, with 90 percent intervals (intended only as a rough guide) estimated under the binomial distribution.

Appendix Figure 4. PIT histogram for GDP inflation
horizon = 1Q



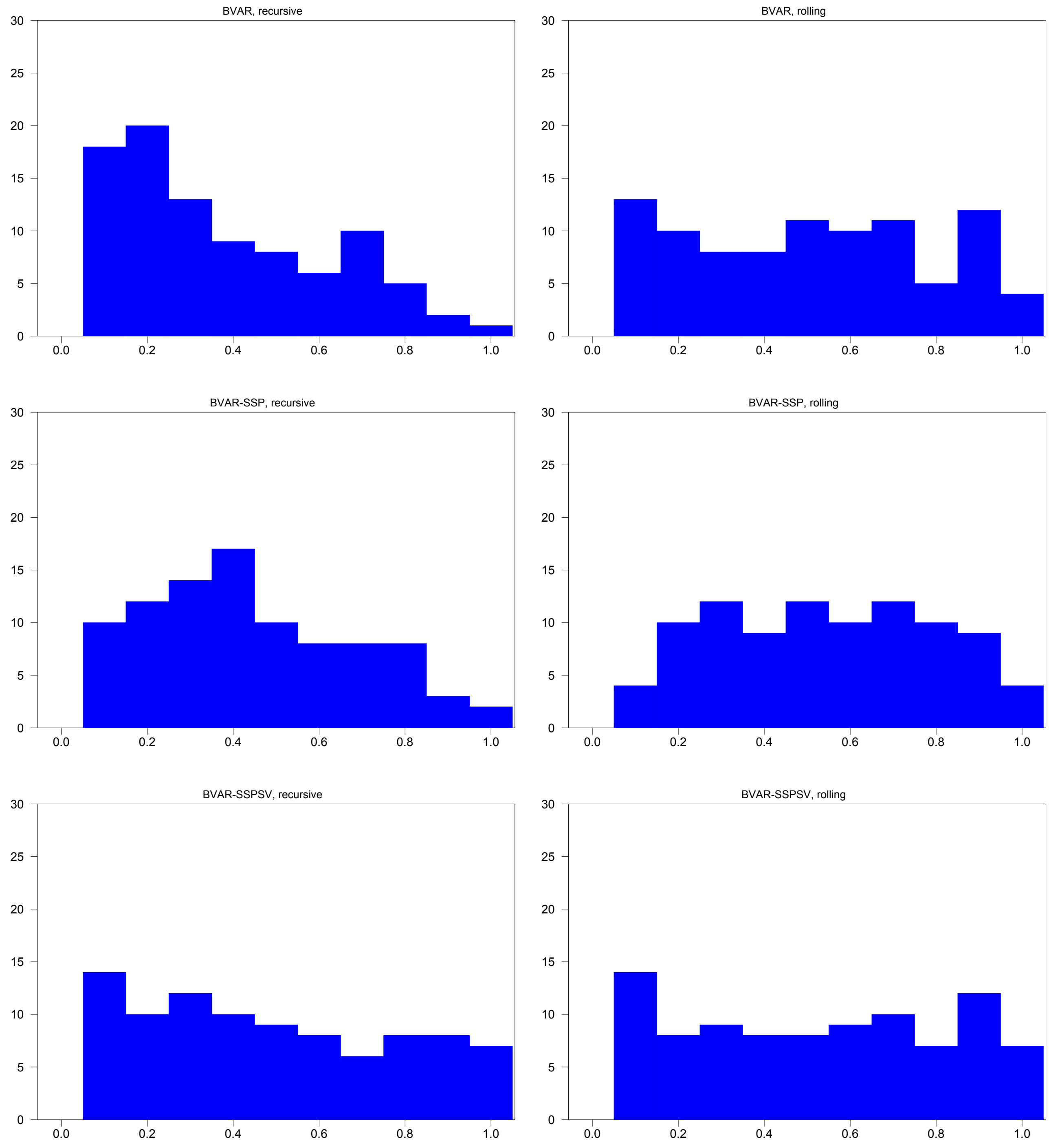
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts, with 90 percent intervals (intended only as a rough guide) estimated under the binomial distribution.

Appendix Figure 5. PIT histogram for Fed funds rate
horizon = 1Q



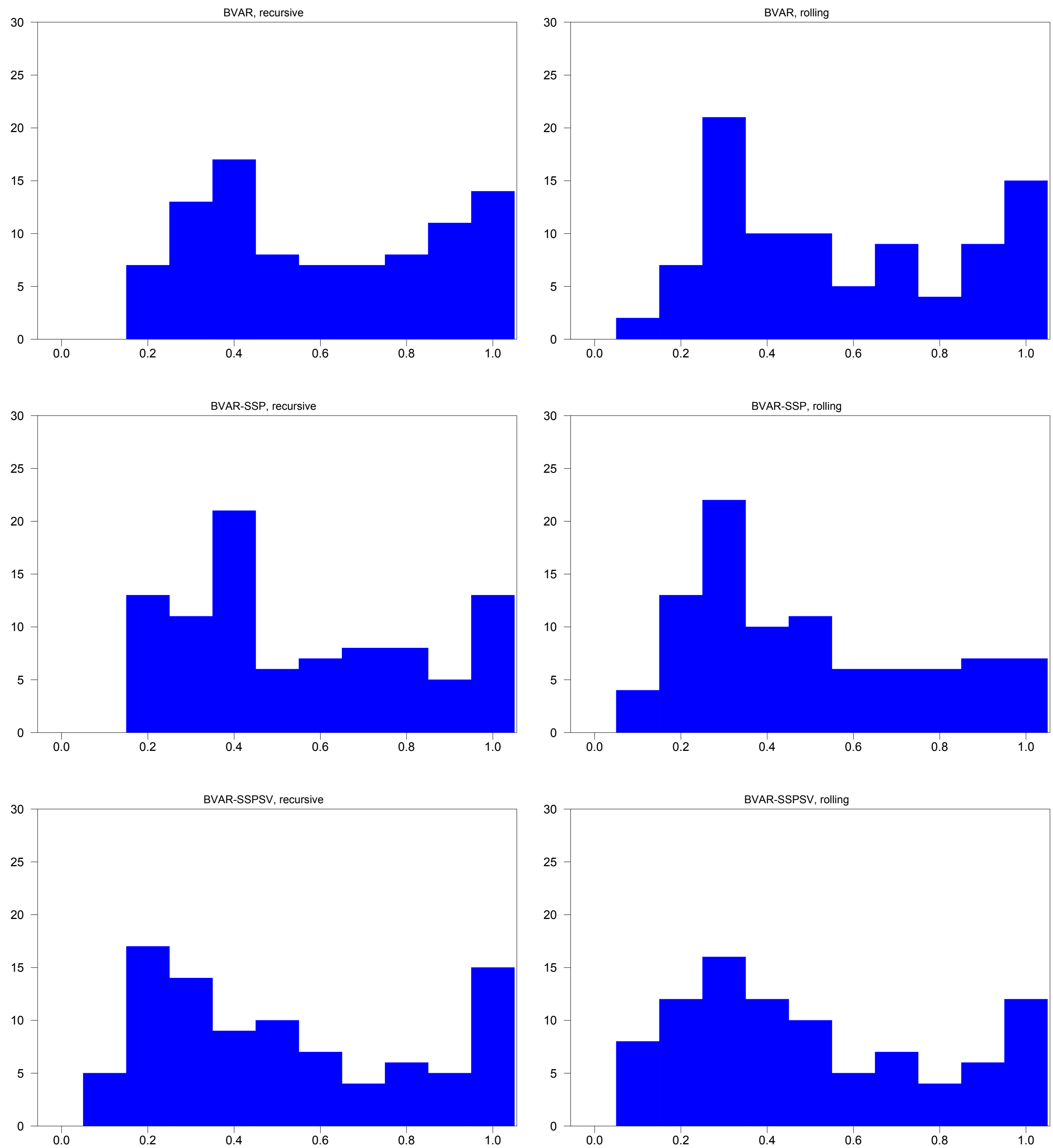
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts, with 90 percent intervals (intended only as a rough guide) estimated under the binomial distribution.

Appendix Figure 6. PIT histogram for GDP growth
horizon = 1Y



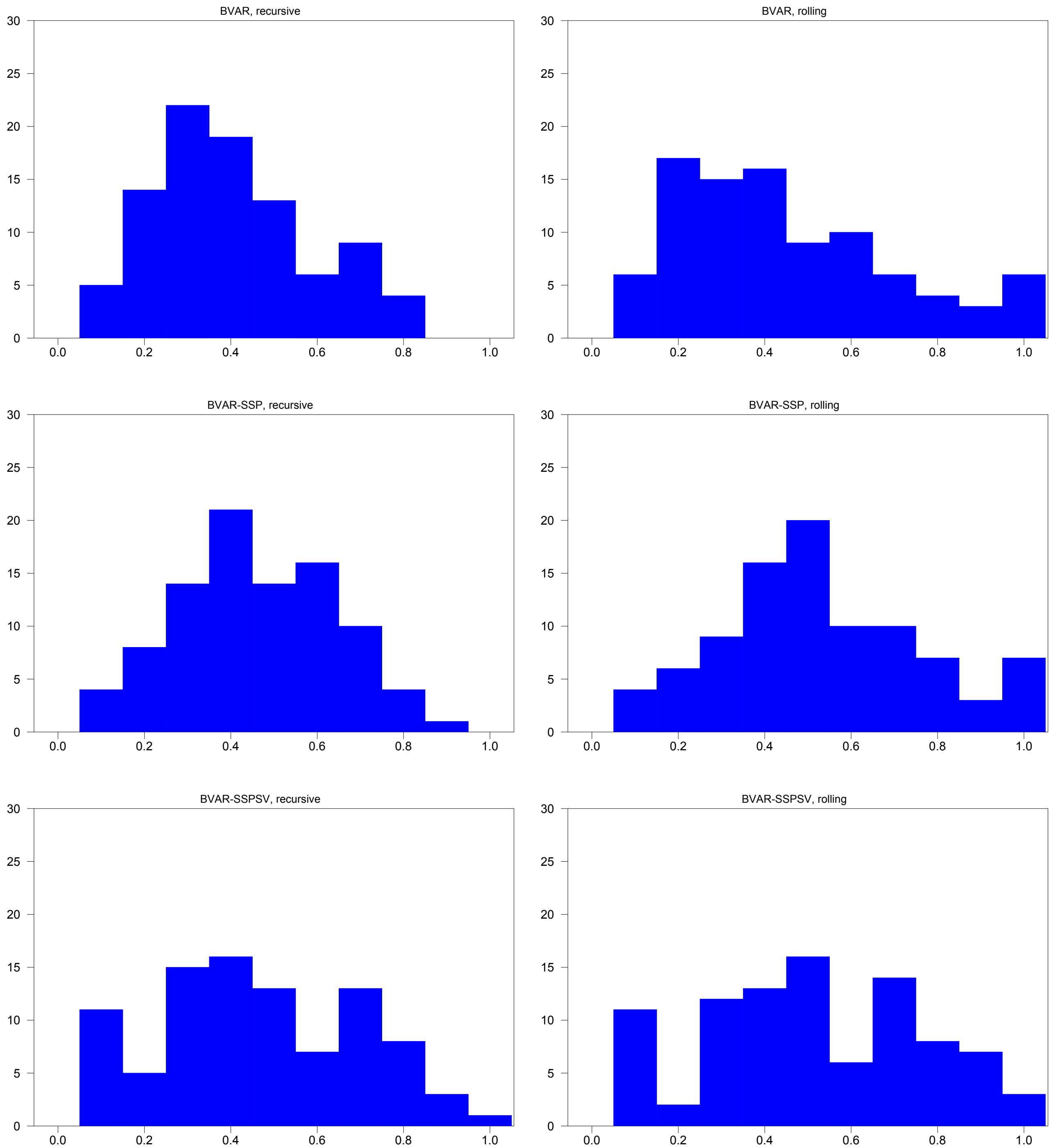
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts.

Appendix Figure 7. PIT histogram for Unemployment
horizon = 1Y



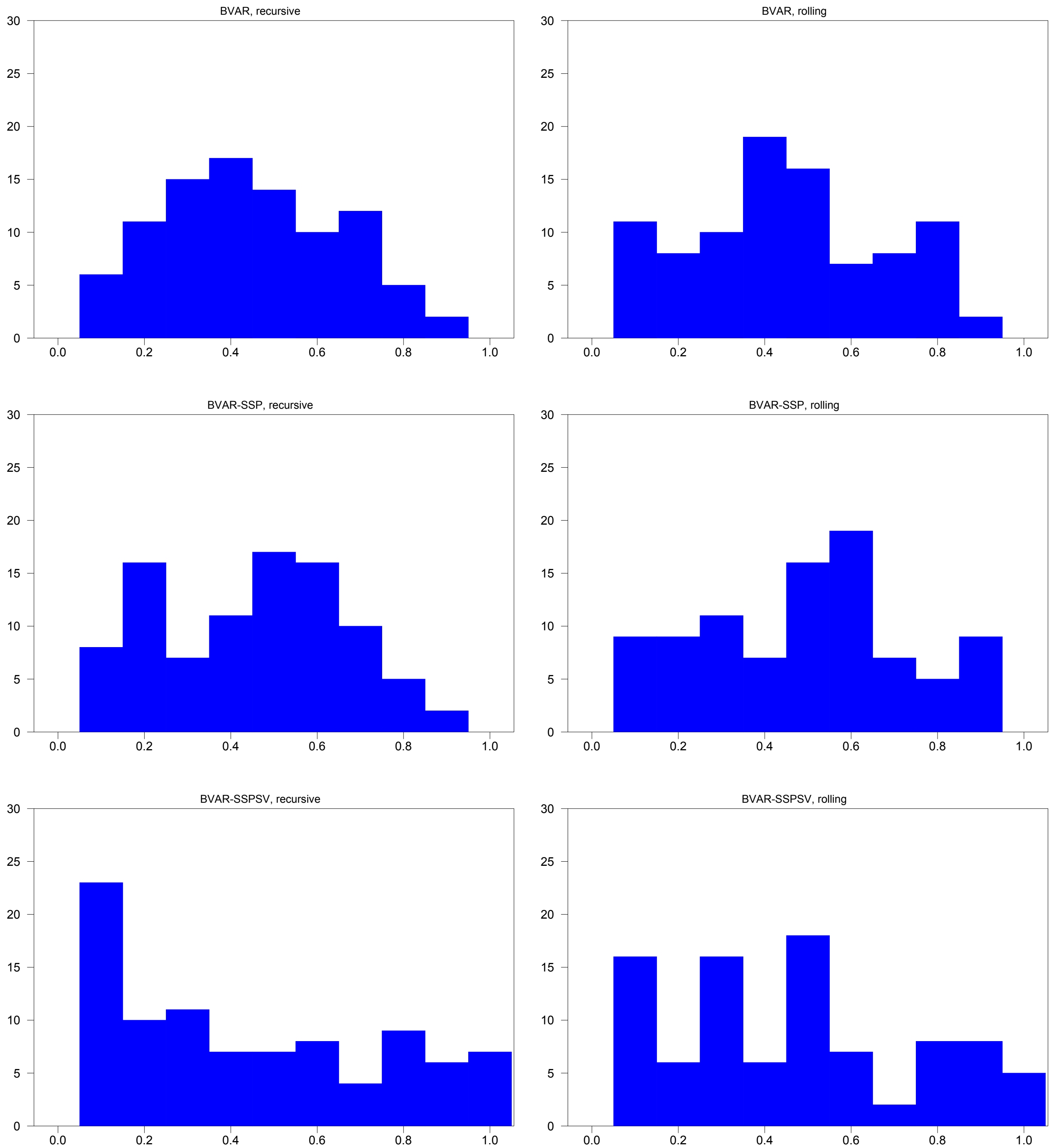
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts.

Appendix Figure 8. PIT histogram for GDP inflation
horizon = 1Y



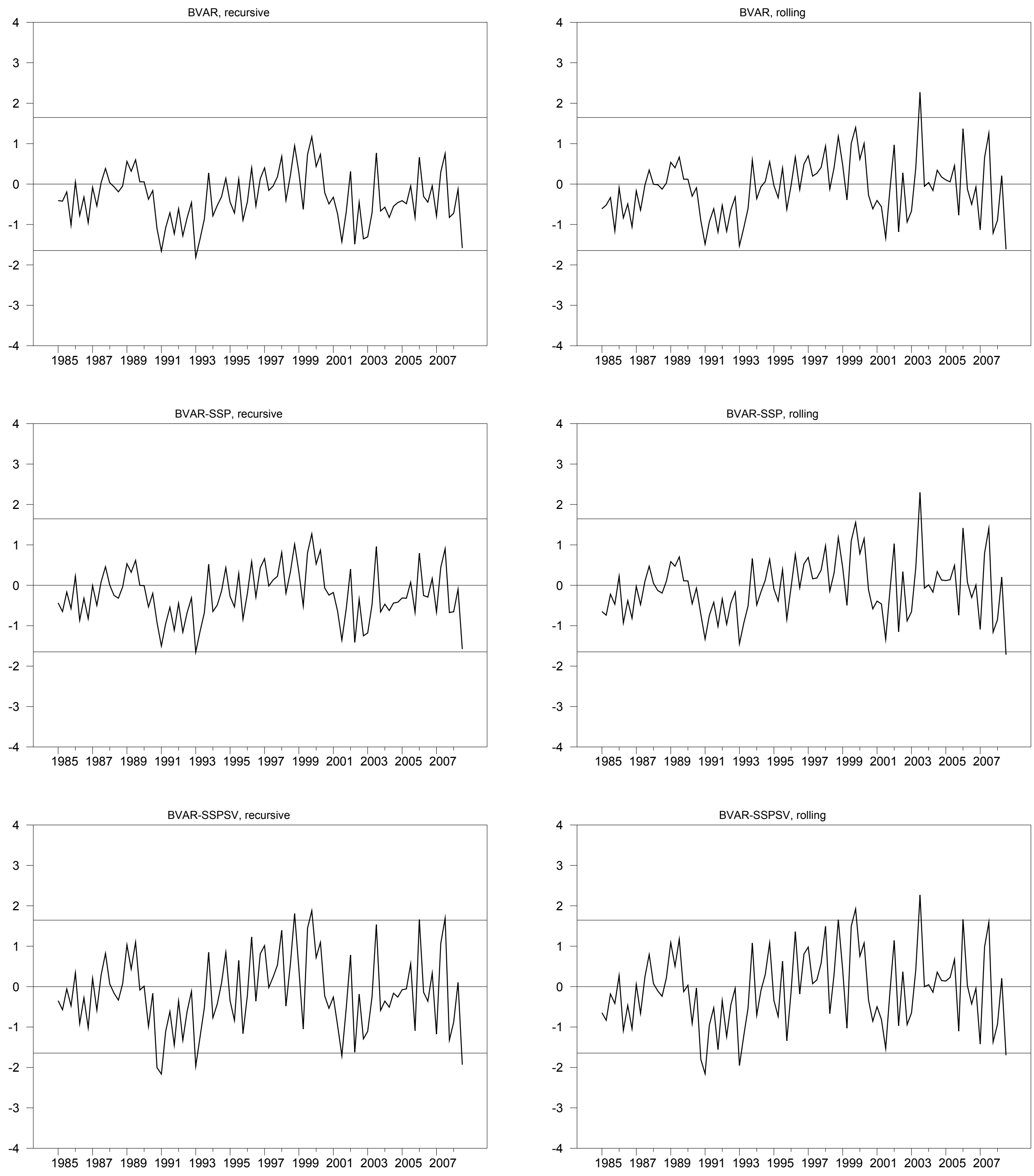
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts.

Appendix Figure 9. PIT histogram for Fed funds rate
horizon = 1Y



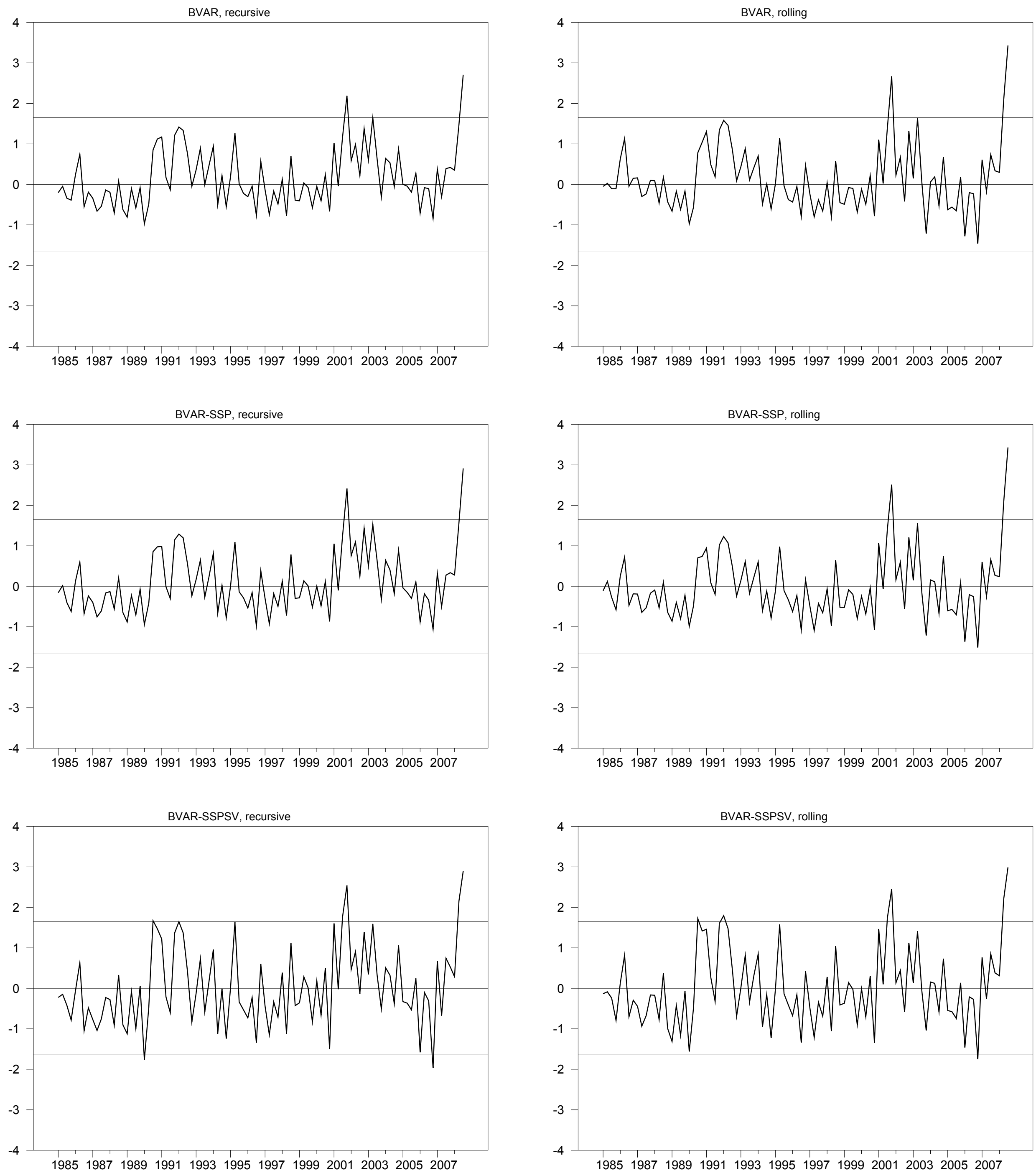
Note: The histograms shown are decile counts of the PITs transforms based on real-time forecasts.

Appendix Figure 10. Normalized Forecast Errors for GDP growth
horizon = 1Q



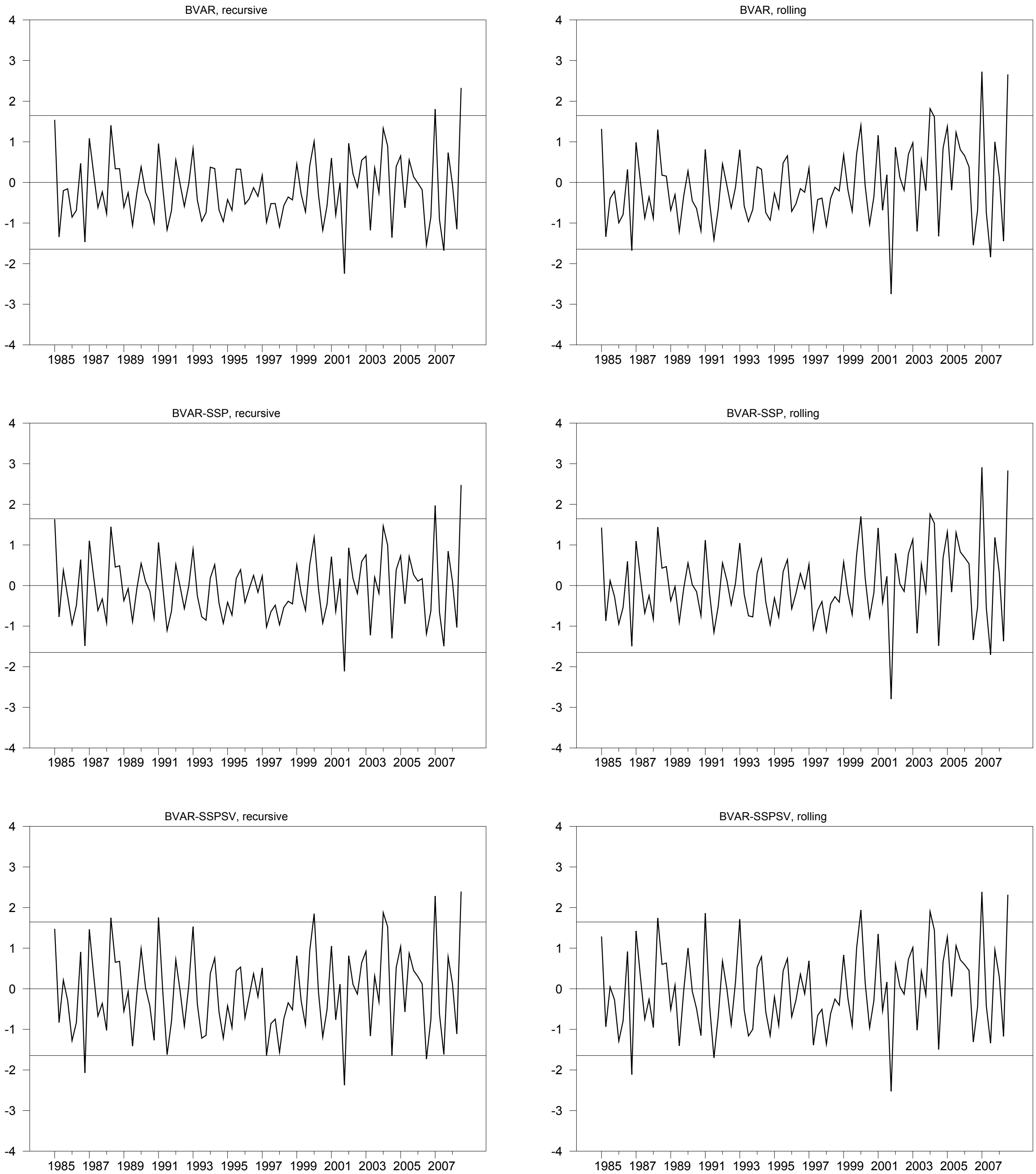
Notes: The normalized forecast errors shown are defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error (generated in real time) and Φ^{-1} is the inverse of the standard normal distribution function. The horizontal lines included in the charts represent 90 percent intervals for the normal distribution.

Appendix Figure 11. Normalized Forecast Errors for Unemployment
horizon = 1Q



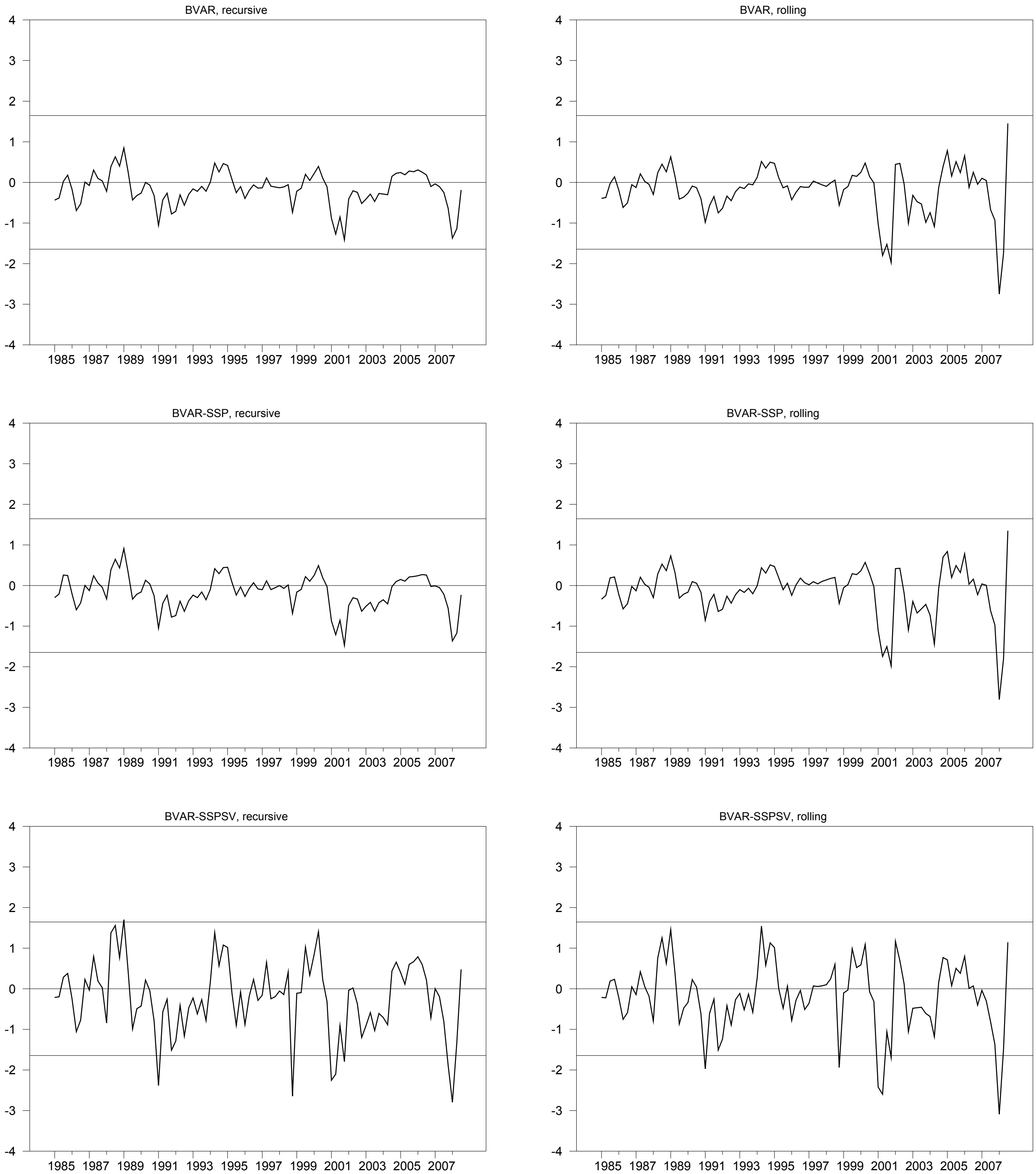
Notes: The normalized forecast errors shown are defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error (generated in real time) and Φ^{-1} is the inverse of the standard normal distribution function. The horizontal lines included in the charts represent 90 percent intervals for the normal distribution.

Appendix Figure 12. Normalized Forecast Errors for GDP inflation
horizon = 1Q



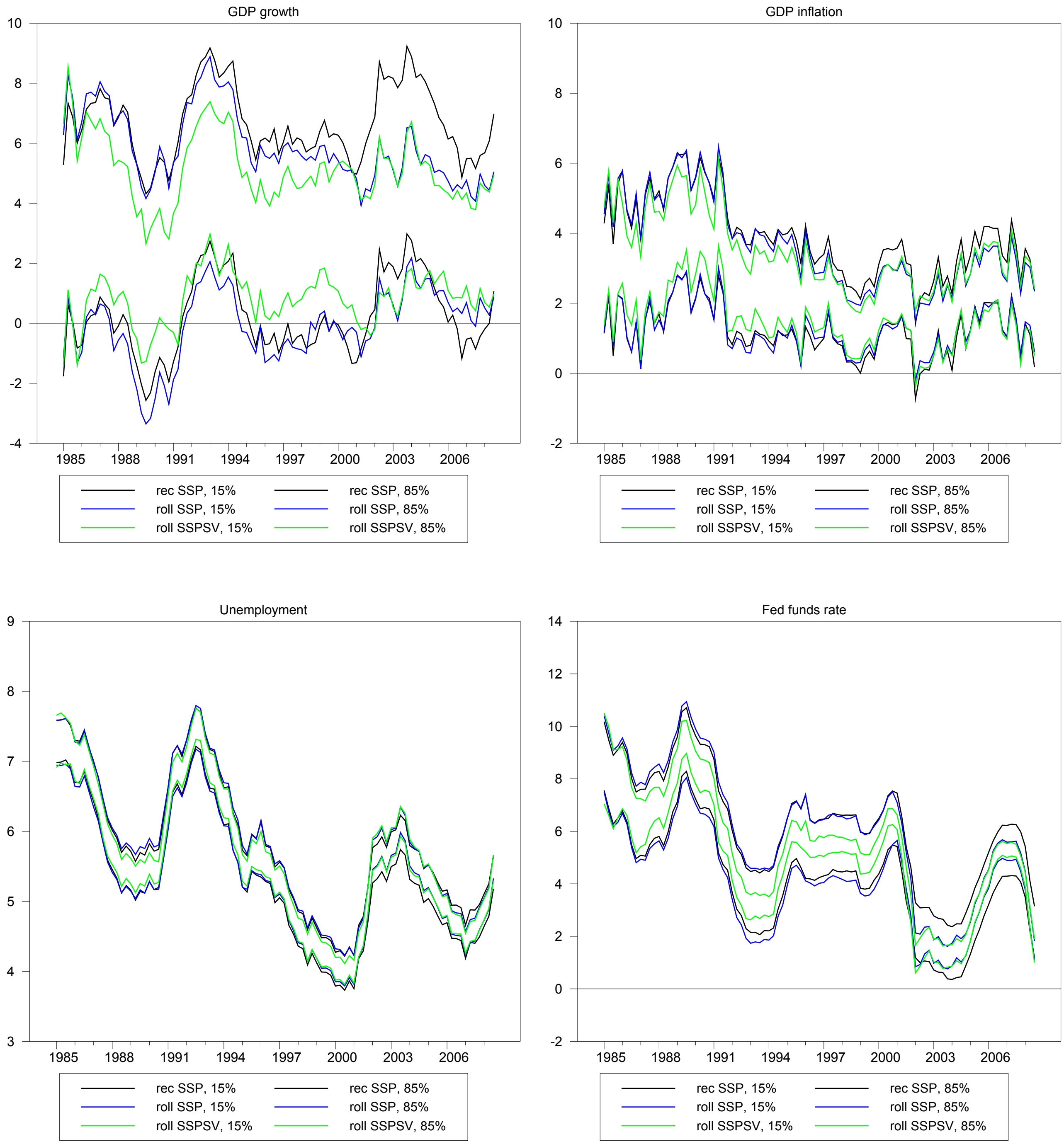
Notes: The normalized forecast errors shown are defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error (generated in real time) and Φ^{-1} is the inverse of the standard normal distribution function. The horizontal lines included in the charts represent 90 percent intervals for the normal distribution.

Appendix Figure 13. Normalized Forecast Errors for Fed funds rate
horizon = 1Q



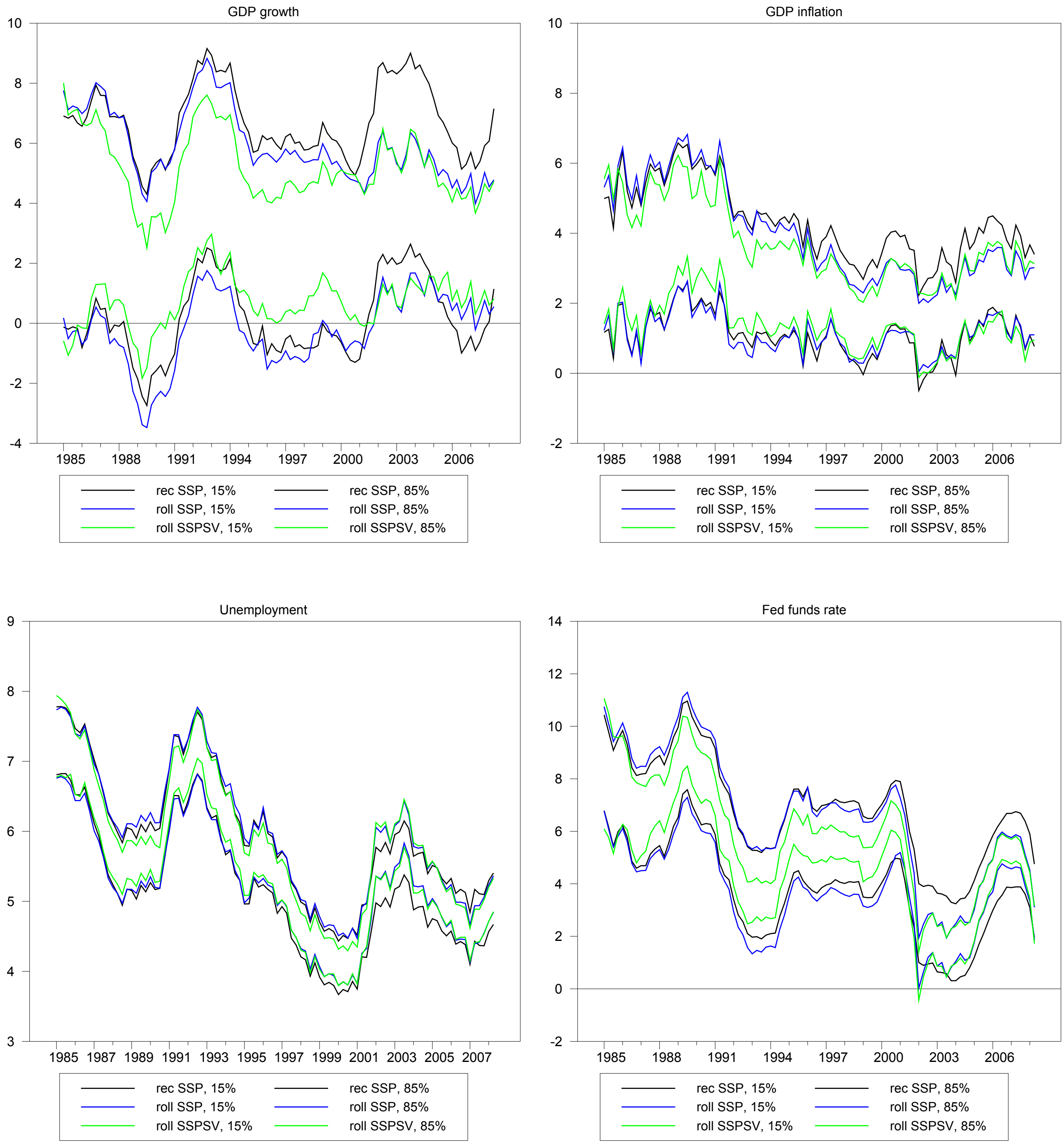
Notes: The normalized forecast errors shown are defined as $\Phi^{-1}(z_{t+1})$, where z_{t+1} denotes the PIT of a one-step ahead forecast error (generated in real time) and Φ^{-1} is the inverse of the standard normal distribution function. The horizontal lines included in the charts represent 90 percent intervals for the normal distribution.

Appendix Figure 14. Time series of 70% intervals
horizon = 1Q



Notes: For each variable, the figure reports time series of 70 percent interval forecasts (percentiles of the marginal density for each variable at the indicated horizon), obtained from the recursive BVAR-SSP (pair of black lines), rolling BVAR-SSP (pair of blue lines), and rolling BVAR-SSPSV (pair of green lines) models.

Appendix Figure 15. Time series of 70% intervals
horizon = 2Q



Notes: For each variable, the figure reports time series of 70 percent interval forecasts (percentiles of the marginal density for each variable at the indicated horizon), obtained from the recursive BVAR-SSP (pair of black lines), rolling BVAR-SSP (pair of blue lines), and rolling BVAR-SSPSV (pair of green lines) models.

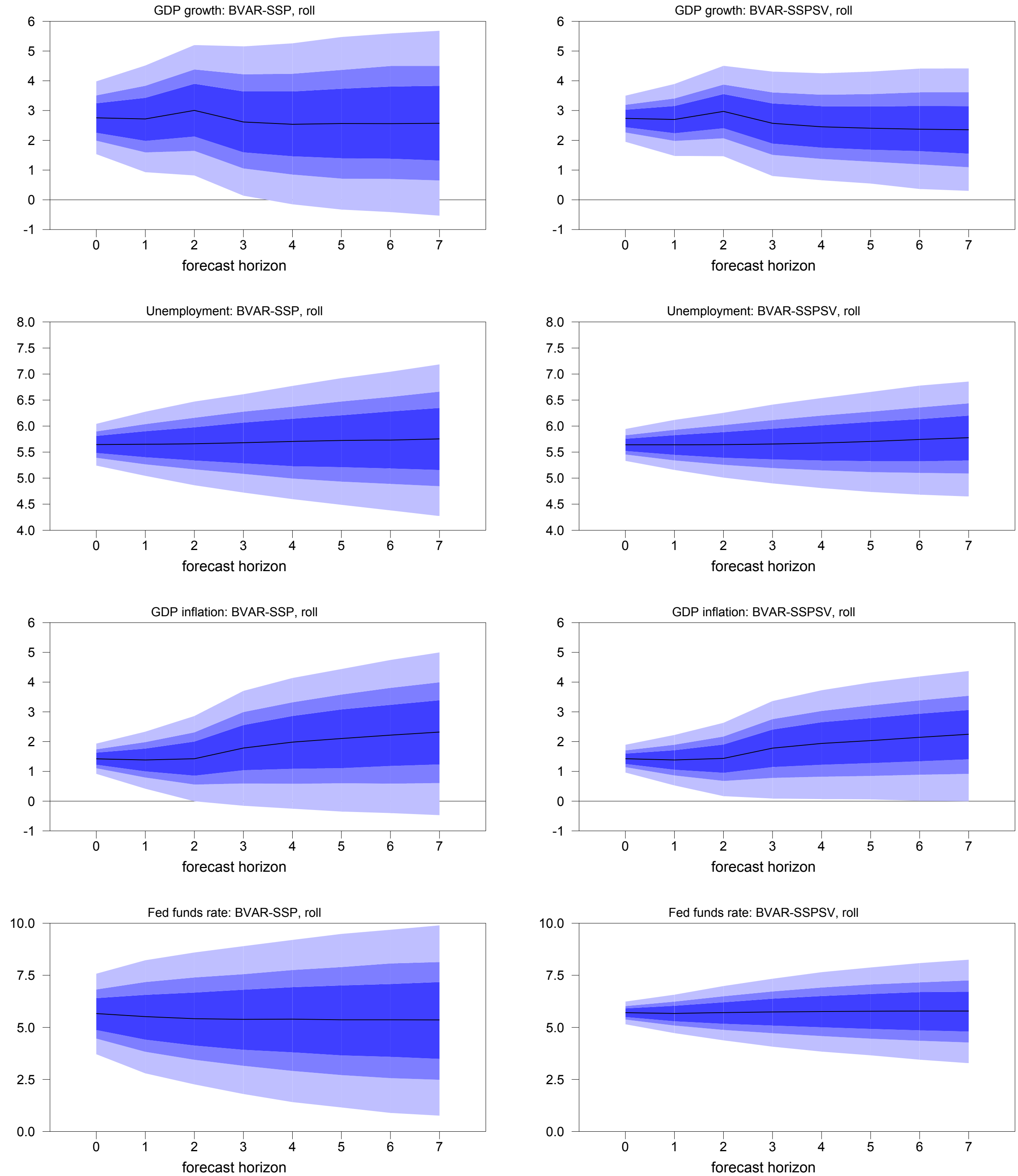
Appendix Figure 16. Time series of 70% intervals
horizon = 1Y

Notes: For each variable, the figure reports time series of 70 percent interval forecasts (percentiles of the marginal density for each variable at the indicated horizon), obtained from the recursive BVAR-SSP (pair of black lines), rolling BVAR-SSP (pair of blue lines), and rolling BVAR-SSPSV (pair of green lines) models.

Appendix Figure 17. Time series of 70% intervals
horizon = 2Y

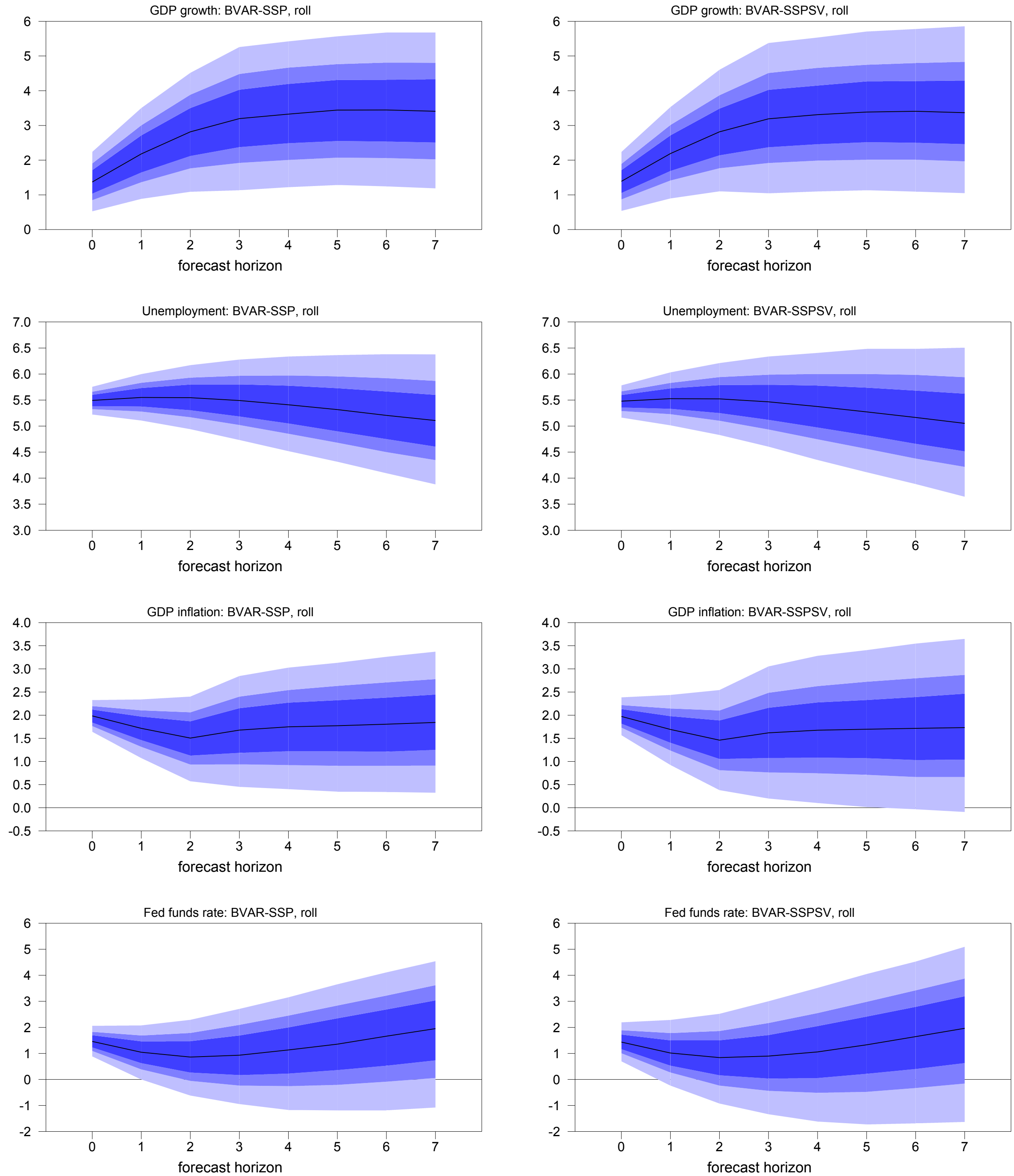
Notes: For each variable, the figure reports time series of 70 percent interval forecasts (percentiles of the marginal density for each variable at the indicated horizon), obtained from the recursive BVAR-SSP (pair of black lines), rolling BVAR-SSP (pair of blue lines), and rolling BVAR-SSPSV (pair of green lines) models.

Appendix Figure 18. Fan charts for forecasts made in 1995:Q4



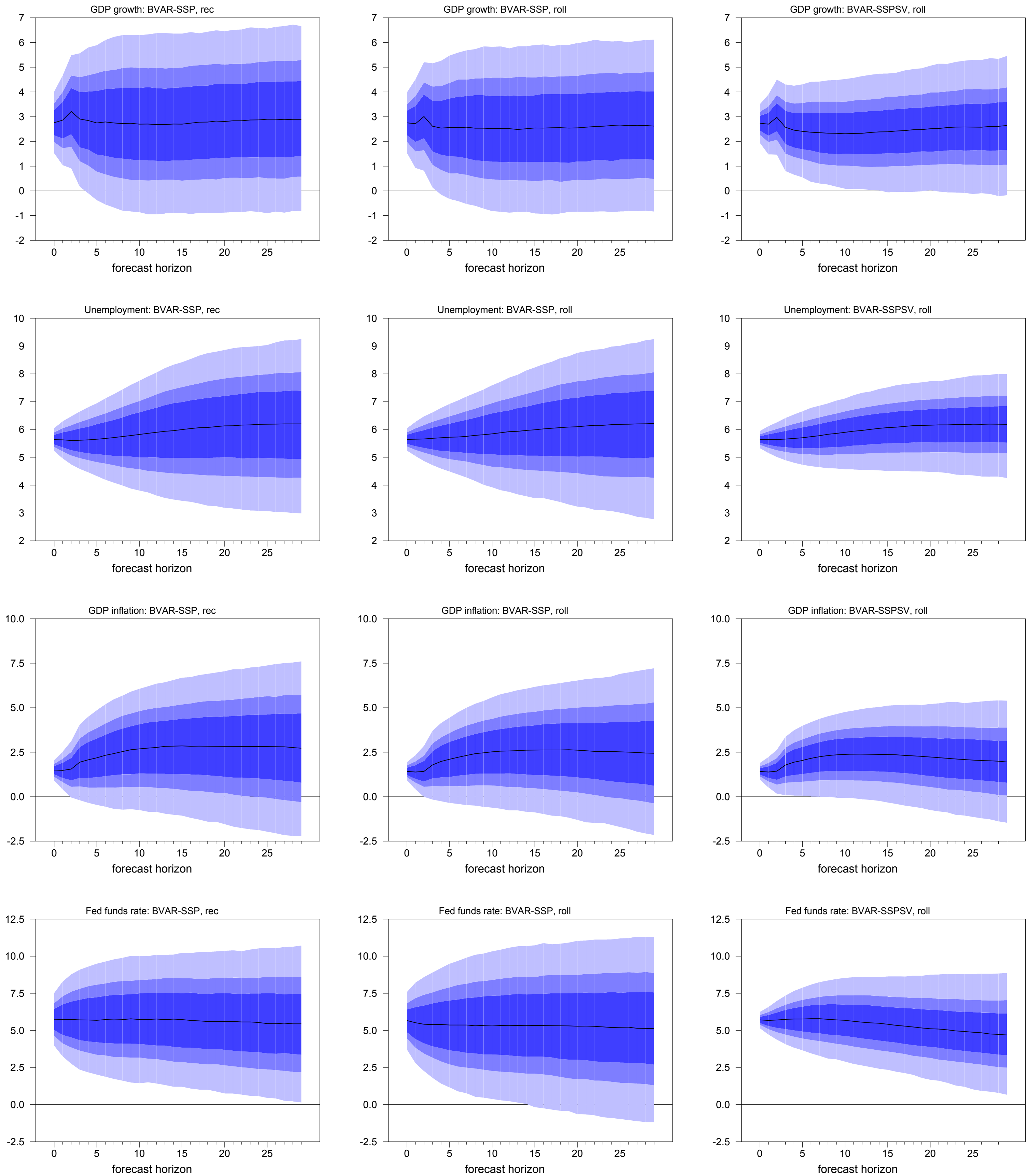
Notes: The figure provides fan charts of forecasts (percentiles of the marginal density for each variable at each horizon) made in the middle of 1995:Q4, using the 1995:Q4 vintage of data, with a data sample ending in 1995:Q3. The GDP growth and inflation forecasts are reported as four-quarter averages. Period 0 on the horizontal axis of each chart refers to the forecast for 1995:Q4. The black line gives the median forecast. The darkest shaded area gives a 50 percent interval. The next shaded area gives a 70 percent interval. The lightest shaded area provides a 90 percent interval.

Appendix Figure 19. Fan charts for forecasts made in 2008:Q3



Notes: The figure provides fan charts of forecasts (percentiles of the marginal density for each variable at each horizon) made in the middle of 2008:Q3, using the 2008:Q3 vintage of data, with a data sample ending in 2008:Q2. The GDP growth and inflation forecasts are reported as four-quarter averages. Period 0 on the horizontal axis of each chart refers to the forecast for 2008:Q3. The black line gives the median forecast. The darkest shaded area gives a 50 percent interval. The next shaded area gives a 70 percent interval. The lightest shaded area provides a 90 percent interval.

Extra Figure 1. Fan charts for long-run forecasts made in 1995:Q4



Extra Figure 2. Fan charts for forecasts made in 2008:Q3