Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors *

Todd E. Clark
Federal Reserve Bank of Cleveland
todd.clark@clev.frb.org

Michael W. McCracken
Federal Reserve Bank of St. Louis
michael.w.mccracken@stls.frb.org

Elmar Mertens
Bank for International Settlements
elmar.mertens@bis.org

This draft: July 6, 2017

Abstract

We develop uncertainty measures for point forecasts from surveys such as the Survey of Professional Forecasters, Blue Chip, or the Federal Open Market Committee's Summary of Economic Projections. At a given point of time, these surveys typically provide forecasts for macroeconomic variables at multiple horizons. To track time-varying uncertainty in the associated forecast errors, we derive a multiple-horizon specification of stochastic volatility. Compared to existing constant-variance approaches, our stochastic-volatility model improves the accuracy of uncertainty measures for survey forecasts.

1   Introduction

A number of central banks use the size of historical forecast errors to quantify the uncertainty around their forecasts. For example, for some years, the Reserve Bank of Australia and the European Central Bank have published forecast fan charts with uncertainty bands derived from historical forecast errors. In the case of the Federal Reserve, since March 2017, the Federal Open Market Committee's (FOMC) Summary of Economic Projections (SEP) has included forecast fan charts with uncertainty bands computed with the root mean square errors (RMSEs) of historical forecasts. (Since 2007, the SEP has included tables of RMSEs of historical forecasts.) As detailed in Reifschneider and Tulip (2007, 2017), the RMSEs are computed from the errors of several different forecasts, including, among others, the Survey of Professional Forecasters (SPF), Blue Chip Consensus, and Congressional Budget Office (CBO). The historical RMSEs are intended to provide an approximate 70 percent confidence interval around the forecast indicated by the median of the FOMC participants' projections.

One important choice central banks must make in such calculations is the sample period of the historical forecast errors. It appears to be commonly recognized that structural changes such as the Great Moderation or unusual periods such as the recent Great Recession can lead to significant shifts in the sizes of forecast errors. For example, in their analysis of historical forecast accuracy (work that underlay the Federal Reserve's initial publication of forecast accuracy measures in the SEP), Reifschneider and Tulip (2007) explicitly chose a sample starting in 1986 to capture accuracy in the period since the start of the Great Moderation. The more recent analysis of Reifschneider and Tulip (2017) discusses some simple evidence of changes in the sizes of forecast errors. In practice, the historical accuracy measures published in the Federal Reserve's SEP are based on a 20-year window of forecast errors. The fan charts of the Bank of England are constructed using information that includes measures of the accuracy over the previous 10 years. Failure to appropriately capture time variation in forecast error variances may result in forecast confidence bands that are either too wide or too narrow, or harm the accuracy of other aspects of density forecasts.

A fairly large literature on the forecast performance of time series or structural models have shown that it is possible to effectively model time variation in forecast error variances.[1]

---

[1]The forecasting literature builds on the initial work of Cogley and Sargent (2005) and Primiceri (2005)

In this work, the historical time variation is large, and modeling it significantly improves the accuracy or calibration of density forecasts. Most such studies have focused on vector autoregressive (VAR) models with stochastic volatility: examples include Carriero, Clark, and Marcellino (2016), Clark (2011), Clark and Ravazzolo (2015), and D'Agostino, Gambetti, and Giannone (2013).[2] Diebold, Schorfheide, and Shin (2016) provide similar evidence for DSGE models with stochastic volatility.

In light of this evidence of time-varying volatility, the accuracy of measures of uncertainty from the historical errors of sources such as SPF, CBO, or the Federal Reserve's Greenbook (known as the Tealbook since mid-2010) might be improved by explicitly modeling their variances as time-varying. Based on the efficacy of stochastic volatility with VAR or DSGE models, a natural starting point might be modeling the available forecast errors as following a stochastic volatility (SV) process. However, the available forecast errors do not immediately fit within the framework of the typical models, because they span multiple forecast horizons, with some correlation or overlap across the horizons. No model exists for this multiple-horizon case, in which the multi-step forecast errors are primitives. (Knuppel (2014) develops an approach for estimating forecast accuracy that accounts for the availability of information across horizons, but under an underlying assumption that forecast error variances are constant over time.) Rather, the typical time series model is specified at a one-step ahead horizon, with multi-step errors inferred from the recursive nature of the parametric model.

Accordingly, in this paper, we develop a multiple-horizon specification of stochastic volatility for forecast errors from sources such as SPF, Blue Chip, the Fed's Greenbook, the FOMC's SEP, or the CBO, for the purpose of improving the accuracy of uncertainty estimates around the forecasts.[3] Our approach can be used to form confidence bands around forecasts that allow for variation over time in the width of the confidence bands; the explicit modeling of time variation of volatility eliminates the need for somewhat arbitrary judgments of sample stability. We focus on forecasts of GDP growth, unemployment, nominal short-term interest rates and inflation from the SPF, including some supplemental results based on forecasts from the Federal Reserve's Greenbook. At each forecast origin, we ob-

---

on VARs with stochastic volatility and Justiniano and Primiceri (2008) on DSGE models with stochastic volatility.

[2]Clark and Ravazzolo (2015) also consider VARs with GARCH and find that VARs with stochastic volatility yield more accurate forecasts.

[3]As will be become clear below, with forecasts such as SPF or Blue Chip, we will work with single forecasts captured by the mean rather than the forecasts of the individual respondents.

serve the forecast error from the previous quarter (measured from advance release data in the case of NIPA variables, as we detail below) and forecasts for the current quarter and the subsequent four quarters. To address the challenge of overlap in forecast errors across horizons, we formulate the model to make use of the current quarter (period t) nowcast error and the forecast updates for subsequent quarters (forecasts made in period t less forecasts made in period $t-1$). These observations reflect the same information as the set of forecast errors for all horizons. However, unlike the vector of forecast errors covering multi-step horizons, the vector containing the forecast updates is serially uncorrelated, under conventional assumptions that SPF forecasts represent a vector of conditional expectations. For this vector of observations, we specify a multiple-horizon stochastic volatility model that can be estimated with Bayesian MCMC methods. From the estimates, we are able to compute the time-varying conditional variance of forecast errors at each horizon of interest.

After developing the model and estimation algorithm, we provide a range of results. First, we estimate the model separately for each variable (growth, unemployment, nominal short-term interest rate, inflation) using the full history of data to document considerable historical variation in forecast error variances, at each forecast horizon. Consistent with evidence from the VAR and DSGE literatures, the forecast error variances shrink significantly with the Great Moderation and tend to rise — temporarily — with each recession, most sharply for the most recent Great Recession. Error variances move together strongly – but not perfectly — across forecast horizons. Second, we produce quasi-real time estimates of forecast uncertainty and evaluate density forecasts implied by the SPF and Greenbook errors and our estimated uncertainty bands. Specifically, we assess forecast coverage rates and the accuracy of density forecasts as measured by the continuous ranked probability score. We show that, by these measures, our proposed approach yields forecasts more accurate than those obtained using simple variances computed with rolling windows of forecast errors as in approaches such as Reifschneider and Tulip (2007, 2017).

Although we focus on estimating forecast uncertainty from models of historical forecast errors featuring time-varying volatility, some survey-based forecasts make available measures of what is commonly termed *ex ante* uncertainty, reflected in forecasts of probability distributions. In the U.S., the one such forecast source is the SPF. However, in the SPF, these probability distributions are provided for just fixed-event forecasts (forecasts for the current and next calendar year) rather than fixed-horizon forecasts, making it difficult to use

the information to compute uncertainty around fixed-horizon forecasts like those available in the point forecasts of SPF most analogous to central bank forecasts commonly published in fan charts. Put another way, making use of the SPF's probability distributions to compare to our main results is hardly feasible (without some very tenuous assumptions necessary to approximate fixed-horizon forecasts from fixed-event forecasts). Moreover, some research has documented some flaws in survey-based probability forecasts, including rounding of responses (e.g., D'Amico and Orphanides 2008 and Boero, Smith, and Wallis 2015) and overstatement of forecast uncertainty at shorter forecast horizons (Clements 2014).[4] Clements (2016) finds that density forecasts obtained from SPF histograms are no more accurate than density forecasts estimated from the historical distributions of past point forecast errors.

Given the vast literature on forecasting, we should emphasize some other choices we have made to constrain the scope of the analysis. The first concerns the distinction between aggregate forecast uncertainty and disagreement across individual forecasters. These concepts are related but distinct (see, e.g., Lahiri and Sheng 2010), and in practice, estimates of the correlations among measures of uncertainty and disagreement vary in the literature. In keeping with the intention of sources such as central bank fan charts, we focus on aggregate forecast uncertainty and leave the direct treatment of disagreement to future research. The second choice concerns the forecasts. In our baseline analysis, we take the forecasts of SPF and Greenbook as given; we do not try to improve them. On this dimension, too, our choice is motivated in part by practices associated with central bank fan charts. For the most part, we leave as a subject for future research the possibility of improving the source forecasts — and in turn our uncertainty estimates — by in some way incorporating additional information from models. However, our extended model that includes a vector autoregressive component is an attempt to allow for possible bias and serial correlation in the expectational updates.

The paper proceeds as follows. Section 2 describes the SPF and Federal Reserve Greenbook forecasts and real time data used in evaluation. Section 3 presents our model of

---

[4]Using data from the ECB's SPF, Abel, et al. (2016) conclude that the squared errors of point forecasts are little correlated with *ex ante* uncertainty obtained from probability distribution forecasts and caution against the use of heteroskedasticity-based measures of uncertainty. However, their comparison uses just squared forecast errors at each moment in time and not more formal, smoother measures of volatility. Moreover, the simple correlation they report does not mean that models of time-varying volatility cannot be used to form reliable confidence intervals around forecasts. In contrast, in an earlier analysis of data from the U.S. SPF, Giordani and Soderlind (2003) find that some GARCH models imply uncertainty estimates that are correlated with *ex ante* uncertainty obtained from probability distribution forecasts.

time-varying variances in data representing multi-horizon forecasts. Section 4 describes our forecast evaluation approach. Section 5 provides results, first on full-sample estimates of volatility and then on various measures of the accuracy of density forecasts. Section 6 concludes.

2   Data

Reflecting in part the professional forecasts available, we focus on quarterly forecasts for a basic set of major macroeconomic aggregates: GDP growth, the unemployment rate, inflation in the GDP price index and CPI, and the 3-month Treasury bill rate.[5] (For simplicity, we use "GDP" and "GDP price index" to refer to output and price series, even though, in our real time data, the measures are based on GNP and a fixed weight deflator for much of the sample.) These variables are commonly included in research on the forecasting performance of models such as VARs or DSGE models. The FOMC's quarterly SEP covers a very similar set of variables, with inflation in the PCE and core PCE price indexes in lieu of the GDP price index or CPI and the federal funds rate in lieu of the T-bill rate. We base most of our results on quarterly forecasts from the SPF, because these forecasts offer two advantages: first, they are publicly available; and second, they offer the longest available quarterly time series of professional forecasts. Alternatives such as Blue Chip are not available publicly or for as long as a sample.[6] However, we also report some supplementary results using forecasts from the Federal Reserve's Greenbook.

We obtained SPF forecasts of growth, unemployment, inflation, and the T-bill rate from the website of the Federal Reserve Bank of Philadelphia. We use SPF forecasts published starting in 1969:Q1 and ending in 2017:Q2 (however, forecasts for CPI inflation and the T-bill start later, and we use data starting in 1981:Q4). At each forecast origin, the available forecasts span five quarterly horizons, from the current quarter through the next four quarters. We form the point forecasts using the mean SPF responses.

We also obtained Greenbook forecasts of growth, unemployment, and inflation from the website of the Federal Reserve Bank of Philadelphia. Although the Federal Reserve prepares forecasts for each FOMC meeting (currently eight meetings per year), we select four forecasts within each year, chosen to align as closely as possible to the timing of the

---

[5]The unemployment rate and T-bill rates are defined as quarterly averages of monthly data. CPI inflation is computed as the percent change in the quarterly average level of the price index.

[6]Reifschneider and Tulip (2007, 2017) find a range of forecast sources, including SPF, Greenbook, and Blue Chip, to have similar accuracy of point forecasts.

SPF forecast published each quarter. We use forecasts published starting in 1966:Q1 and ending in 2011:Q4 (however, forecasts for CPI inflation do not begin until 1980:Q1). The end of the sample reflects the five year delay in the Federal Reserve's public release of the forecasts. We are unable to include T-bill rate forecasts from Greenbook because they are not included in the Philadelphia Fed's data files.[7] At each forecast origin, we include forecasts spanning five quarterly horizons, from the current quarter through the next four quarters.

Quantifying the forecast errors underlying our analysis requires a choice of outcomes against which to measure the forecasts.[8] To form accurate confidence bands around the forecast (and density forecasts more generally) at the time the forecast is produced, in roughly the middle of quarter t, we measure the quarter $t-1$ forecast error with the first (in time) estimate of the outcome. Specifically, for real GNP/GDP and the associated price deflator, we obtain real-time measures for quarter $t-1$ data as it was publicly available in quarter t from the quarterly files of real-time data compiled by the Federal Reserve Bank of Philadelphia's Real Time Data Set for Macroeconomists (RTDSM). 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. Because revisions to quarterly data for the unemployment rate, CPI inflation, and the T-bill rate are relatively small or non-existent in the case of the T-bill rate, we simply use the currently available data to measure the outcomes and corresponding forecast errors.[9] We obtained data on the unemployment rate, CPI, and 3-month Treasury bill rate from the FRED database of the Federal Reserve Bank of St. Louis.

Before we turn from data to our model, note that as a general matter, our model can be readily applied to forecasts from other sources. As the introduction notes, the forecasts need to be of the fixed horizon type (not fixed event) and cover (in sequence) multiple forecast horizons. The forecasts can be at any data frequency, although quarterly would be most typical in macroeconomic settings. Although our data on growth and inflation are quarter-on-quarter percent changes, our model could be applied to use year-on-year percent

---

[7]Studies such as Faust and Wright (2008) and Reifschneider and Tulip (2017) make use of short-term interest rate forecasts from Greenbook obtained from the Federal Reserve's Board of Governors. However, as discussed in Faust and Wright (2008, 2009), for much of the available history, these forecasts have been tied to conditioning assumptions about monetary policy, rather than unconditional forecasts. Accordingly, we do not include interest rates in our Greenbook assessment.

[8]Sources such as Romer and Romer (2000), Sims (2002), and Croushore (2006) discuss various considerations for assessing the accuracy of real-time forecasts.

[9]For evidence on CPI revisions, see Kozicki and Hoffman (2004).

changes.[10]

3 Model

To set the stage for our multivariate analysis, we first review the implications of a simple, standard autoregressive model with stochastic volatility. We then turn to the more complex setting of the forecasts available to us and the model we consider in this paper. We conclude by describing a constant variance benchmark included in the empirical analysis.

3.1 Example of standard AR-SV specification

In standard time series models — univariate or multivariate — allowing time-variation in forecast uncertainty has become common and straightforward.[11] For example, a simple time series model for a scalar variable y_t could take the form of an AR(1) specification with stochastic volatility:

$$
\begin{aligned}
y_t &= by_{t-1} + \lambda_t^{0.5}\epsilon_t, \ \epsilon_t \sim N(0,1) \\
\log(\lambda_t) &= \log(\lambda_{t-1}) + \nu_t, \ \nu_t \sim N(0,\phi).
\end{aligned}
$$

In this case, at the population level, the h-step ahead prediction error (from a forecast origin of period t) is given by

$$
e_{t+h} = \lambda_{t+h}^{0.5}\epsilon_{t+h} + b\lambda_{t+h-1}^{0.5}\epsilon_{t+h-1} + \cdots + b^{h-1}\lambda_{t+1}^{0.5}\epsilon_{t+1},
$$

with forecast error variance

$$
\begin{aligned}
E_t e_{t+h}^2 &= E_t(\lambda_{t+h} + b^2\lambda_{t+h-1} + \cdots + b^{2(h-1)}\lambda_{t+1}) \\
&= \lambda_t \sum_{j=0}^{h-1} b^{2j} \exp\left(\frac{1}{2}(h-j)\phi\right).
\end{aligned}
$$

Forecast uncertainty at all horizons is time-varying due to the stochastic volatility process, given by the random walk model for $\log(\lambda_t)$. In practice, for such a model, forecast uncertainty is normally estimated using simulations of the posterior distribution of forecasts, which involve simulating future realizations of volatility, shocks, and y paths. Note, however, that these simulations key off the single process for y_t and the single process for $\log(\lambda_t)$. Contrary to the observed data on survey forecasts, such a model would thus imply

---

[10] In this case, the primary changes would relate to the specifics of the aggregation matrix polynomial $\mathbf{B}(L)$ described below.

[11] See, for example, Clark (2011) and D'Agostino, Gambetti, and Giannone (2013).

that the levels as well as the volatilities of forecast errors are perfectly correlated across forecast horizons.

3.2   Our model for multi-horizon forecasts

Accommodating time-variation in forecast uncertainty associated with forecasts such as SPF or Blue Chip (or the FOMC's SEP) is more complicated than in the standard autoregressive model with stochastic volatility applied to time series data. In this section we make clear why and our solution to the complication.

3.2.1   Forecast error decomposition

We assume a data environment that closely reflects the one we actually face with SPF forecasts (the same applies with forecasts from sources such as Blue Chip and the Federal Reserve's Greenbook). At any given forecast origin t, we observe forecasts of a scalar variable y_t. Reflecting data availability, the previous quarter's outcome, y_{t-1}, is known to the forecaster, and we assume the current-quarter outcome y_t is unknown to the forecaster. For simplicity, we define the forecast horizon h as the number of calendar time periods relative to period t, and we denote the longest forecast horizon available as H. We describe the forecast for period $t + h$ as an h-step ahead forecast, although outcomes for period t are not yet known. The SPF compiled at quarter t provides forecasts for $t + h$, where $h = 0, 1, 2, 3, 4$, and $H = 4$, such that, at each forecast horizon, we have available $H + 1$ forecasts (the nowcast for period t and forecasts for periods 1 through H).

In practice, exactly how the forecast is constructed is unknown, except that the forecast likely includes some subjective judgment and does not come from a simple time series model. We will treat the point forecast as the conditional expectation $E_t y_{t+h}$; at the forecast origin t, we observe the forecasts $E_t y_t$, $E_t y_{t+1}$, ..., $E_t y_{t+H}$, as well as the forecasts made in previous periods. Reflecting real-time data timing, the conditioning information underlying the expectation does not include the actual value of y_t. We seek to estimate forecast uncertainty defined as the conditional variance, $\text{var}_t(y_{t+h})$, allowing the forecast uncertainty to be time-varying.

The challenge in this environment is in accounting for possible overlapping information in the multi-step forecasts (or forecast errors) observed at each forecast horizon. Knuppel (2014) develops an approach for estimating forecast accuracy that accounts for such overlap in observed forecast errors, but under the implicit assumption that forecast error variances

are constant over time.

To model time variation in forecast uncertainty in overlapping forecasts, we make use of a decomposition of the multi-step forecast error into a nowcast error and the sum of changes (from the previous period to the current period) in forecasts for subsequent periods. To simplify notation, let a subscript on the left-side of a variable refer to the period in which the expectation is formed and a subscript on the right side refer to the period of observation. So $_t y_{t+h}$ refers to the h-step ahead expectation of y_{t+h} formed at t, and $_t e_{t+h}$ refers to the corresponding forecast error. We will refer to the error $_{t+h} e_{t+h}$ — the error in predicting period $t+h$ from an origin of period $t+h$ without known outcomes for the period — as the nowcast error. Denote the forecast updates — which we will refer to as expectational updates — as $\mu_{t+h|t} \equiv {}_t y_{t+h} - {}_{t-1} y_{t+h} = (E_t - E_{t-1}) y_{t+h}$.

The starting point of our decomposition is an accounting identity, which makes the h-step ahead forecast error equal the sum of (i) the error in the nowcast that will be formed h steps ahead and (2) a sequence of expectational updates that occur between the current period through the next h periods for the expected value at $t+h$:[12]

$$_t e_{t+h} = {}_{t+h} e_{t+h} + \sum_{i=1}^{h} \mu_{t+h|t+i}, \ \forall \ h \geq 1. \tag{1}$$

To see the basis of this relationship, consider a simple example of a two-step ahead forecast error. We obtain the relationship by starting from the usual expression for the two-step error and then adding and subtracting forecasts from the right side as follows:

$$
\begin{aligned}
_t e_{t+2} &= y_{t+2} - {}_t y_{t+2} \\
&= (y_{t+2} - {}_{t+2} y_{t+2}) + ({}_{t+2} y_{t+2} - {}_t y_{t+2}) \\
&= (y_{t+2} - {}_{t+2} y_{t+2}) + ({}_{t+2} y_{t+2} - {}_{t+1} y_{t+2}) + ({}_{t+1} y_{t+2} - {}_t y_{t+2}) \\
&= {}_{t+2} e_{t+2} + \mu_{t+2|t+2} + \mu_{t+2|t+1}.
\end{aligned}
$$

Note that, in this decomposition, the information structure of real-time forecasts from a source such as SPF — in which, as noted above, forecasts made at time t reflect information that does not yet include knowledge of the realized value of y_t — adds a term to the decomposition that would not exist with textbook setups of time series models in which

---

[12]Some previous studies have also made use of expectational updates, for different purposes. For example, Patton and Timmermann (2012) write a short-horizon forecast as a sum of a long-horizon forecast and forecast revisions and use it as the basis of an optimal revision regression to test forecast optimality.

forecasts made at t reflect information through t.[13]

To obtain our baseline econometric framework, we proceed to embed some basic expectational restrictions. By construction, the expectational update $\mu_{t+h|t}$ forms a martingale difference sequence:

$$E_{t-1} \, \mu_{t+h|t} = 0. \tag{2}$$

Assuming that, at every forecast origin t, the forecast source (SPF or Greenbook) provides us with a vector of conditional expectations, it then follows from (2) that the terms in (1) are uncorrelated with each other. As detailed below, we will exploit this in our econometric model and in our (Bayesian) simulation of the posterior distribution of forecast errors, from which we are able to compute the uncertainty around multi-step forecasts using the decomposition (1) with uncorrelated terms.

As we detail below, we use Bayesian methods to measure forecast uncertainty as captured in the posterior distribution of forecast errors. Our treatment can be seen as reflecting the following variance analytics. Under the martingale difference assumption on the expectational updates, we can characterize the conditional variance of the multi-step forecast error as the sum of the (expected) variances of the individual terms in (1):

$$\mathrm{var}_t(y_{t+h}) = \mathrm{var}_t({}_te_{t+h}) = \mathrm{var}_t({}_{t+h}e_{t+h}) + \sum_{i=1}^{h} \mathrm{var}_t(\mu_{t+h|t+i})$$

$$= E_t\left[\mathrm{var}_{t+h}({}_{t+h}e_{t+h})\right] + \sum_{i=2}^{h} E_t\left[\mathrm{var}_{t+i-1}(\mu_{t+h|t+i})\right] + \mathrm{var}_t(\mu_{t+h|t+1}). \tag{3}$$

The simplification from the first line of (3) to the second uses the law of total variance, which implies:

$$\mathrm{var}_t(\mu_{t+h|t+i}) = E_t\left[\mathrm{var}_{t+i-1}(\mu_{t+h|t+i})\right] + \mathrm{var}_t\left[\underbrace{E_{t+i-1}(\mu_{t+h|t+i})}_{=0}\right], \tag{4}$$

where the last term collapses to zero because of the martingale difference property of $\mu_{t+h|t+i}$; a similar argument holds for the conditional variance of the future nowcast error in (3).

Although we quantify forecast uncertainty from simulations of the posterior predictive distribution with an approach detailed below, this decomposition could be used to build up

---

[13]With a standard time series model setup, in which it is typically assumed that $E_t y_t = y_t$, the accounting identity would contain h components, taking the same form as in (1) but without the first term (which equals 0 in the standard time series setup): ${}_te_{t+h} = \sum_{i=1}^{h} \mu_{t+h|t+i}$.

estimates of $\text{var}_t({}_te_{t+h})$ from estimates of the conditional variances, for (1) the variance of the nowcast error, $\text{var}_t({}_te_t)$, and (2) the variance of the expectational update of forecasts for horizon $i = 1, \ldots, h$, $\text{var}_t(\mu_{t+i|t+1})$. Note that these are exactly as many variances as we have observables. The martingale difference property of updates to the survey expectations provides an orthogonalization of the data that, conditional on knowing the variances of expectational updates, obviates the need to estimate correlations.

3.2.2 Model of time-varying volatility

Based on the decomposition (1) and the martingale difference assumption (2), we specify a multivariate stochastic volatility model for the available nowcast error and expectational updates. As noted above, the forecast origin (denoted t) is roughly the middle of quarter t, corresponding to the publication of the survey forecast. At the time the forecasters construct their projections, they have data on quarter $t-1$ and some macroeconomic data on quarter t. We construct a data vector strictly contained in that information set and define the data vector to contain $H+1$ elements: the nowcast error for quarter $t-1$ and the revisions in forecasts for outcomes in quarters t through $t+H-1$. (Although at origin t the forecasts go through period $t+H$, the available forecast revisions only go through period $t+H-1$.) In the case of the SPF, which publishes forecasts for the current and next four quarters, corresponding to $H = 4$ in our notation, we have the nowcast error and four forecast updates to use. For comparability, our analysis of Greenbook forecasts relies on the same choice of horizons.[14]

More specifically, we define the data vector as:

$$
\boldsymbol{\eta_t} = \begin{bmatrix} y_{t-1} - E_{t-1}y_{t-1} \\ (E_t - E_{t-1})y_t \\ (E_t - E_{t-1})y_{t+1} \\ \vdots \\ (E_t - E_{t-1})y_{t+H-1} \end{bmatrix} = \begin{bmatrix} {}_{t-1}e_{t-1} \\ \mu_{t|t} \\ \mu_{t+1|t} \\ \vdots \\ \mu_{t+H-1|t\cdot} \end{bmatrix} \tag{5}
$$

This specification includes an offset in timing between the first element of $\boldsymbol{\eta_t}$ and the remaining elements, by pairing the $t-1$ nowcast error — the most recently observed nowcast error at the forecast origin t — with the t updates in expectations. The offset is consistent with the deliberate construction of $\boldsymbol{\eta_t}$ as a martingale difference sequence relative to E_{t-1}

---

[14]The maximum forecast horizon available in Greenbook fluctuates over the course of a calendar year and varies over the course of the data history. Accommodating longer forecast horizons than the $H = 4$ we use would mean accommodating more missing values.

and with the publication of actual data. Based on the accounting identity (1), given the vector $\boldsymbol{\eta_t}$, we are able to obtain the forecast errors from:

$$\mathbf{e_t} = \begin{bmatrix} {}_t e_t \\ \vdots \\ {}_{t-h} e_t \end{bmatrix} = \mathbf{B}(L)\boldsymbol{\eta_{t+1}}, \tag{6}$$

where $\mathbf{B}(L)$ is a known lag polynomial containing zeros and ones.

Our baseline model of the expectational updates is a multivariate stochastic volatility specification, allowing for correlation both across elements of $\boldsymbol{\eta}$ as well as across innovations to the log volatilities of each component of $\boldsymbol{\eta_t}$. In other words we allow for correlation between both level and scale shocks to $\boldsymbol{\eta}$:

$$
\begin{aligned}
\boldsymbol{\eta_t} &= \mathbf{A}\tilde{\boldsymbol{\eta}}_{\mathbf{t}} \qquad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ a_{H+1,1} & a_{H+1,2} & \dots & & 1 \end{bmatrix} \\
\tilde{\boldsymbol{\eta}}_{\mathbf{t}} &= \boldsymbol{\Lambda_t}^{0.5}\boldsymbol{\epsilon_t}, \ \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \mathbf{I_{H+1}}), \ \boldsymbol{\Lambda_t} \equiv \operatorname{diag}(\lambda_{1,t}, \dots, \lambda_{H+1,t}) \\
\log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t}, \quad i = 1, \dots, H+1 \\
\boldsymbol{\nu_t} &\equiv (\nu_{1,t}, \nu_{2,t}, \dots, \nu_{H+1,t})' \sim N(\mathbf{0}, \boldsymbol{\Phi}).
\end{aligned}
\tag{7}
$$

where $\mathbf{A}$, a lower triangular matrix with values of 1 on the diagonal, serves to capture correlations across the components of $\boldsymbol{\eta_t}$ while correlations across the innovations to stochastic volatility are captured by $\boldsymbol{\Phi}$. The variance-covariance matrix of $\boldsymbol{\eta_t}$ is given by $\boldsymbol{\Sigma_t} = \mathbf{A}\boldsymbol{\Lambda_t}\mathbf{A}'$.

While measures of correlation between elements of $\boldsymbol{\eta}$ do not enter directly in the variance calculus laid out above, the inclusion of non-zero lower-triangular coefficients in $\mathbf{A}$ matters, at least somewhat, for our estimates, since we need to resort to full-information, Bayesian sampling methods to estimate the time-varying volatilities as explained further below. Moreover, some non-zero correlation between elements of $\boldsymbol{\eta_t}$ should generally be expected, as persistence in the underlying macroeconomic variables forecasted by the SPF should lead survey respondents to jointly revise updates in expectations of a given variable at different horizons. In fact, if SPF forecasts were generated from the simple, univariate AR-SV specification described above, expectational updates contained in $\boldsymbol{\eta_t}$ would be perfectly correlated with each other. For similar reasons, we allow innovations to log volatilities to be correlated across the components of $\boldsymbol{\eta_t}$, following the multivariate volatility specification of studies such as Primiceri (2005). We obtained similar results for a model treating the volatility innovations as mutually independent (as in, e.g., Cogley and Sargent 2005).

As this specification suggests, our focus in this paper is on a model of time-varying volatility. For that purpose, we build the model around forecast errors and expectational updates that are assumed mean zero.[15] Reifschneider and Tulip (2017) also assume future forecasts to be unbiased, treating any past historical bias as transitory.

By choosing an otherwise conventional, conditionally linear and Gaussian data-generating process, our approach will yield prediction intervals and densities that are symmetric (and not skewed). In doing so, we follow the broader literature (see references above) on including stochastic volatility in time series models for macroeconomic forecasting. The last subsection of the results section discusses possible extensions to accommodate asymmetries.

For readers concerned that the baseline model abstracts from asymmetries, our model can still be used to reliably estimate the conditional forecast error variance. Specifically, our approach can be used to obtain estimates of second moments. As long as these sample moments are accurate measures (consistent estimates) of the population moments, and as long as the η_t's are martingale differences, we are able to consistently estimate forecast error variances — even with non-normal, non-symmetric distributions. These variances would provide a useful measure of forecast uncertainty even if the underlying forecast distributions featured some asymmetries.

Although the baseline model features symmetry, the observed forecast errors and expectational updates need not be Gaussian. The model makes use of conditional innovations (in ϵ_t) that are Gaussian, but this does not imply the observed forecast errors and expectational updates to be Gaussian. In fact, the model implies that the distributions of the observed expectational updates and forecast errors feature fat tails. We discuss below a model extension to treat the conditional innovations ϵ_t as having fat tails.

3.2.3 Generalized model without MDS assumption

As noted above, our baseline specification reflects an assumption that the vector of expectational updates forms a martingale difference sequence, consistent with full rationality of the forecasts. This assumption helps to yield a parsimonious model, and parsimony is well known to be helpful in forecasting. However, studies such as Croushore (2010) and Reifschneider and Tulip (2017) provide evidence of some biases in forecasts from sources such as SPF and Greenbook. Moreover, recent research by Coibion and Gorodnichenko (2015)

---

[15]In preliminary analysis, we obtained similar results when, before estimating the model, we demeaned the elements of the data vector η_t using a quasi-real time approach to computing a time-varying mean, with one-sided exponential smoothing.

and Mertens and Nason (2015), among others, has shown survey-based forecasts to display information rigidities, reflected in some serial correlation in forecast errors.

To allow for possible biases and persistence in forecast errors and expectational updates, we also consider an extension of our model that does not rest on the MDS assumption. In this case, we make use of the accounting identity (1) that relates forecast errors to expectational updates, but we not impose the MDS assumption underlying our baseline model. Specifically, we consider a vector autoregressive (VAR) model of the expectational updates with stochastic volatility, taking the same form introduced in Cogley and Sargent (2005) that has by now been considered in a number of forecasting studies:

$$\boldsymbol{\eta_t} = \mathbf{C_0} + \mathbf{C_1}\boldsymbol{\eta_{t-1}} + \mathbf{A}\boldsymbol{\Lambda_t}^{0.5}\boldsymbol{\epsilon_t}, \tag{8}$$

where $\mathbf{C_0}$ is a vector of intercepts, $\mathbf{C_1}$ is a matrix of slope coefficients, and the remainder of the model is defined as in the baseline specification. Although the model could easily be extended to include longer lags, we have deliberately chosen to include just one lag, for parsimony and the likely low-order serial correlation in the expectational updates. As a simple check of the extent of the correlation, to get a sense of the serial correlation in the expectational updates, for each variable we estimated vector autoregressions based on the vector $\boldsymbol{\eta_t}$ using 0 to 4 lags and assessed fit with the BIC. The BIC indicates the optimal lag order to be 0 for the unemployment rate and CPI inflation and 1 for GDP growth, GDP inflation, and the T-bill rate. Our use of a (Bayesian) VAR with one lag appears consistent with this simple check.

As detailed below, we estimate this extended model — referred to below as the VAR-SV specification — with conventional Minnesota-type priors on $\mathbf{C_0}$ and $\mathbf{C_1}$. This model allows for non-zero means and serial correlation of the expectational updates. We obtain the forecast errors using the accounting identity $\mathbf{e_t} = \mathbf{B}(L)\boldsymbol{\eta_{t+1}}$, where $\mathbf{B}(L)$ is a known lag polynomial containing zeros and ones.

3.2.4 Estimating the model and forecast uncertainty

The baseline model of (7) and the extension (8) can be estimated by Bayesian Markov chain Monte Carlo methods (a Gibbs sampler). We focus on describing the estimation of the baseline specification; the estimation of the VAR model involves adding a conventional Gibbs step to draw the VAR coefficients from their conditional posterior (see, e.g., Clark

and Ravazzolo 2015).[16] The baseline model's algorithm involves iterating over the following three blocks: First, taking estimates of $\mathbf{\Lambda_t}^{0.5}$ as given, we employ recursive Bayesian regressions with diffuse priors to estimate the lower triangular coefficients of $\mathbf{A}$, which is tantamount to a Choleski decomposition of $\boldsymbol{\eta}$ into $\tilde{\boldsymbol{\eta}}$.[17] Second, we estimate the stochastic volatilities of $\tilde{\boldsymbol{\eta}}_{\mathbf{t}}$ using the multivariate version of the Kim, Shephard, and Chib (1998) [henceforth, KSC] algorithm introduced into macroeconomics by Primiceri (2005). Third, given draws for the sequences of $\log(\lambda_{i,t})$ for all i and t we estimate the variance-covariance matrix of innovations to the SV processes, Φ, using an inverse Wishart conjugate-prior centered around a mean equal to a diagonal matrix with 0.2^2 on its diagonal using $10 + H$ degrees of freedom, which makes the prior slightly informative. Note that our setting of the prior mean is in line with settings used in some studies of stochastic volatility, including Stock and Watson (2007) and Clark (2011).

For notational generality, let $\tilde{\eta}_{i,t}$ refer to the i-th element of $\tilde{\boldsymbol{\eta}}_{\mathbf{t}}$. Squaring and taking the log of $\tilde{\eta}_{i,t}$ yields the measurement equation

$$\log \tilde{\eta}_{i,t}^2 = \log \lambda_{i,t} + \log \epsilon_{i,t}^2, \ \ i = 1, \ldots, H+1, \tag{9}$$

with corresponding transition equation

$$\log(\lambda_{i,t}) = \log(\lambda_{i,t-1}) + \nu_{i,t}, \ \ i = 1, \ldots, H+1.$$

As the different elements of $\tilde{\boldsymbol{\eta}}_{\mathbf{t}}$ (as opposed to $\boldsymbol{\eta}_{\mathbf{t}}$) are mutually uncorrelated, the measurement equation (9) includes a χ^2-distributed innovation that is independent across i. KSC develop a mixture-of-normals approximation to its distribution. Accordingly, the state space representation and simulation smoother of Durbin and Koopman (2002) can be used to estimate the model. The state space representation also allows us to easily handle the occurrences of a few missing observations in SPF and Greenbook forecasts in our sample. KSC and Primiceri (2005) provide additional detail on the estimation algorithm.[18]

To estimate the uncertainty around multi-step forecasts, we simulate the posterior distribution of forecast errors using the model (7) and an approach like that detailed in Cogley,

---

[16]For the VAR's coefficients, the prior means are all zero, and the standard deviations take the Minnesota form, with the hyperparameter governing overall shrinkage set at 0.2, the hyperparameter for "other" lags relative to "own" lags set at 0.5, and the hyperparameter governing intercept shrinkage set at 1.

[17]To initialize the Gibbs sampler, we draw initial values for each lower-triangular coefficients of $\mathbf{A}$ from independent standard normal distributions.

[18]Our implementation of the algorithm incorporates the correction of Primiceri (2005) developed in Del Negro and Primiceri (2015).

Morozov, and Sargent (2005). For each forecast horizon h, we need to simulate draws of the forecast error $_te_{t+h}$, which is the sum of uncorrelated terms given in equation (1). We obtain draws of these terms by simulating forward the vector $\boldsymbol{\eta_t}$ of our multivariate SV model, to obtain the posterior distribution of forecast errors.

We obtain these draws with the following steps, for each draw of parameters of the MCMC algorithm. Note that, to evaluate forecasts for horizons up to H steps ahead, the timing of the data and model involves simulating $H + 1$ periods at each forecast origin.

1. For each component i of $\boldsymbol{\tilde{\eta}_t}$, simulate $\log \lambda_{i,t}$ forward from period $t+1$ through period $t + H + 1$ using its random walk process and its shock, obtained by simulating the vector of shocks with variance-covariance matrix $\boldsymbol{\Phi}$.

2. Simulate the time path of $N(\mathbf{0}, \mathbf{I_{H+1}})$ innovations $\boldsymbol{\epsilon_t}$ forward from period $t+1$ through period $t + H + 1$.

3. Obtain the time path of $\boldsymbol{\tilde{\eta}_{t+h}}$ from period $t + 1$ through period $t + H + 1$ as the product of the simulated $\boldsymbol{\Lambda_{t+h}}^{0.5}$ and $\boldsymbol{\epsilon_{t+h}}$.

4. Transform $\boldsymbol{\tilde{\eta}_t}$ into $\boldsymbol{\eta_t}$ by multiplication with $\mathbf{A}$.

5. At each horizon h, construct the draw of the forecast error by summing the relevant terms from the previous step according to the decomposition (1). Construct the draw of the forecast by adding the forecast error to the corresponding point forecast from SPF.

Given the set of draws produced by this algorithm, we compute the forecast statistics of interest. For example, we compute 68 percent coverage rates as the percentages of the observed forecast errors that fall within the 68 percent bands (credible sets) of the simulated forecast error distributions. In the next section, we detail these and the other forecast evaluation metrics considered.

3.2.5 An alternative approach assuming constant variances of forecast errors

In light of common central bank practice (e.g., Reifschneider and Tulip 2007, 2017 and the fan charts in the Federal Reserve's SEP), the most natural benchmark against which to compare our proposed model-based approach is one based on historical forecast error variances treated as constant over some window of time and into the future. That is, at

each forecast origin t, prediction intervals and forecast densities can be computed assuming normally distributed forecast errors with variance equal to the variance of historical forecast errors over the most recent R periods (e.g., the SEP fan charts are based on forecast errors collected over the previous 20 calendar years).[19] Accordingly, we report results obtained under a similar approach, where we will collect continuously updated estimates generated from rolling windows of forecast errors covering the most recent $R = 60$ quarterly observations. For simplicity, below we will refer to this specification as the "constant variance" approach and denote it with "FE-CONST," even though it acknowledges the potential for variance changes over time by using a rolling window of observations.[20] Note, too, that this benchmark approach differs from our model-based approach in that the benchmark uses forecast errors directly, whereas our model-based approach uses the expectational updates and obtains forecast errors as linear combinations of the expectational updates.

Of course, a key choice is the size of the rolling window (R) used in the constant variance approach. As noted above, some central banks use windows of 40 or 80 quarterly observations; Clements (2016) uses 50 quarterly observations. In our analysis, there is an important sample tradeoff in data availability: making the rolling window bigger shortens the forecast sample available for evaluation. Accordingly, in our baseline results, we essentially split the difference, so to speak, and use a rolling window of 60 observations in the constant variance benchmark. With this setting, we have available the following samples for the evaluation of SPF forecasts: 1984:Q1–2017:Q1 for GDP growth, unemployment, and GDP inflation; and 1996:Q4–2017:Q1 for CPI inflation and the T-bill rate. While the sample of available Greenbook forecasts permits similar start dates, the end date for evaluating is 2011:Q4, reflecting the five-year blackout period for publication. As we detail in the robustness results below, our main findings apply to rolling windows shorter or longer than the baseline.

4   Evaluation metrics

The previous section described two alternative volatility models — our proposed stochastic volatility model, our extension to a VAR with stochastic volatility, and a constant-variance benchmark. This section considers two measures of density forecast accuracy to assess the

---

[19]A number of other studies, such as Kenny, Kostka, and Masera (2014) and Rossi and Sekhposyan (2014), have also used normal distributions based on a given point forecast and error variance.

[20]We obtained very similar results with a more parametric approach or model assuming a time-invariant normal distribution for the nowcast error and the expectational updates collected in $\boldsymbol{\eta_t}$ (while maintaining the martingale difference sequence assumption): $\boldsymbol{\eta_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. We employ Bayesian methods to estimate this model within the quasi-real-time setup described above, assuming a diffuse inverse-Wishart prior.

absolute and relative performance of these models. The first measure focuses on the accuracy of prediction intervals. 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) and Clark (2011) have used interval forecasts as a measure of the calibration of macroeconomic density forecasts. Accordingly, we will report the frequency with which real-time outcomes for growth, unemployment, inflation, and the Treasury bill rate fall inside 68 percent prediction intervals, as well as some selected empirical quantiles of the forecast distribution.[21] A frequency of more (less) than 68 percent means that, on average over a given sample, the estimated forecast density is too wide (narrow). We will judge the significance of the results on the basis of p-values for the null of correct coverage (empirical = nominal rate), based on t-statistics.[22]

Our second measure of density accuracy is the continuous ranked probability score (CRPS). As indicated in Gneiting and Raftery (2007) and Gneiting and Ranjan (2011), some researchers view the CRPS as having advantages over the log score.[23] In particular, the CRPS does a better job of rewarding values from the predictive density that are close to but not equal to the outcome, and it is less sensitive to outlier outcomes. The CRPS, defined such that a lower number is a better score, is given by

$$CRPS_t(y_{t+h}^o) = \int_{-\infty}^{\infty} \left(F(z) - 1\{y_{t+h}^o \leq z\} \right)^2 dz = E_f|Y_{t+h} - y_{t+h}^o| - 0.5 E_f|Y_{t+h} - Y_{t+h}'|,$$
(10)

where F denotes the cumulative distribution function associated with the predictive density f, $1\{y_{t+h}^o \leq z\}$ denotes an indicator function taking value 1 if the outcome $y_{t+h}^o \leq z$ and 0 otherwise, and Y_{t+h} and Y_{t+h}' are independent random draws from the posterior predictive density. We compute the CRPS using the empirical CDF-based approximation given in equation (10) of Krueger, et al. (2017).

As noted above, a number of studies have compared the density forecast performance

---

[21]We also considered analyzing results for the 90 percent prediction interval. However, with more observations available at the 68 percent level than the 90 percent level, we chose to devote attention to the former rather than the latter.

[22]The t-statistics are based on HAC variances computed with the Newey and West (1987) estimator and a lag order equal to the forecast horizon plus 2. These p-values should be viewed 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 the forecast densities considered in this paper are obtained from estimated models or at least variances.

[23]Recent applications can be found, for example, in Ravazzolo and Vahey (2014) and Clark and Ravazzolo (2015).

of time series models with stochastic volatility against time series models with constant variances (e.g., Clark 2011, Clark and Ravazzolo 2015, D'Agostino, Gambetti, and Giannone 2013, and Diebold, Schorfheide, and Shin 2016). In some cases, the models with constant variances are estimated with rolling windows of data. In some respects, the comparisons in this paper are similar to these studies. However, a key difference is that we take the point forecasts as given from a source such as the SPF, whereas, in these papers, the point forecasts vary with each model. So, for example, in Clark's (2011) comparison of a BVAR with stochastic volatility against a BVAR with constant variances estimated over a rolling window of data, the use of a rolling window affects the point forecasts. As a result, the evidence on density forecast accuracy from the VAR and DSGE literature commingles effects of conditional means and variances with rolling windows versus other estimators and other models. In contrast, in this paper, by using point forecasts from SPF or Greenbook, we are isolating influences on density accuracy due to variances.

5   Results

We begin this section of results with a brief review of the data properties and with full-sample estimates of stochastic volatility. We then provide the out-of-sample forecast results, first on coverage and then on density accuracy as measured with the CRPS. The next subsection provides a summary of various robustness checks, including results for Greenbook forecasts. The section concludes with a discussion of directions in which the model could be extended.

5.1   Full-sample

As noted above, the data used to estimate our model are the expectational updates (for simplicity, defined broadly here to include the nowcast error) contained in $\boldsymbol{\eta_t}$. Figures 1 and 2 report these data for GDP growth and the unemployment rate, respectively; the data for the other variables, along with the forecast errors, are provided in the supplementary appendix, in the interest of brevity. Qualitatively, the results we highlight for GDP growth and unemployment also apply to the other variables.

As implied by the forecast error decomposition underlying our model, the expectational updates are fairly noisy. Although there is some small to modest serial correlation in the data on the longer-horizon expectational updates, this serial correlation is much smaller

than that in the multi-step forecast errors. As an example, for the unemployment rate, compare the 4-step ahead expectational update in Figure 2 to the 4-step ahead forecast errors in the appendix Figure 4.

Another notable feature of the data is that, at longer forecast horizons, the expectational updates are smaller in absolute size than are the corresponding forecast errors. This feature is more or less inherent to expectational updates. In addition, in most cases (less clearly so for the unemployment rate than the other variables), the absolute sizes of the expectational updates appear to be larger in the period before the mid-1980s than afterward, consistent with the Great Moderation widely documented in other studies. For growth, unemployment, and the T-bill rate, the expectational errors tend to be larger (in absolute value) in recessions than expansions.

Figures 3 to 7 provide the time-varying volatility estimates obtained with the expectational updates. Specifically, the red lines in each figure provide the full-sample (smoothed) estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating an historical volatility path at each forecast origin; these estimates underlay the forecast results considered in the next section. Note that, to improve chart readability by limiting the number of panels on each page to four, we omit from each chart the estimates for the 3-step ahead forecast horizon; these unreported estimates are consistent with the results summarized below.

Across variables, the volatility estimates display several broad features, as follows.

- The time variation in volatility is considerable. The highs in the volatility estimates are typically 3 to 4 times the levels of the lows in the estimates.

- Some of the time variation occurs at low frequencies, chiefly with the Great Moderation of the 1980s. The Great Moderation is most evident for GDP growth, the unemployment rate (less so for the nowcast horizon than longer horizons), and inflation in the GDP price index. For CPI inflation, the volatility estimate declines even though the available sample cuts off most of the period preceding the typical dating of the Great Moderation. For the T-bill rate, for which the sample is shorter, as with

20

the CPI, the SV estimate shows a sharp falloff at the beginning of the sample; this falloff is consistent with SV estimates from time series models obtained with longer samples of data (e.g., Clark and Ravazzolo 2015).

- Some of the time variation is cyclical, as volatility has some tendency to rise temporarily around recessions. For example, the volatility of GDP growth and unemployment rises with most recessions, and the volatility of the T-bill rate picks up around the 2001 and 2007-2009 recessions. The cyclical pattern appears smaller for inflation, except that CPI inflation spiked sharply around the time of the Great Recession, presumably due to the dramatic, unexpected falloff in inflation that occurred as commodity prices collapsed.

- The overall magnitude of volatility for the nowcast horizon versus the expectational updates for longer horizons varies by variable, probably reflecting data timing. For growth and both measures of inflation, the level of volatility at the nowcast horizon exceeds the level of volatility at longer horizons. However, for the unemployment rate and T-bill rate, nowcast volatility is lower than longer-horizon update volatility, probably because the nowcast is often or always formed with the benefit of one month of data (for the unemployment rate and T-bill rate) on the quarter.

- For the most part, for the period since the 1980s, the contours of SV estimates for inflation in the GDP price index and CPI are similar. There are of course some differences, including the relatively sharp late-2000's rise for the CPI that probably reflects a bigger influence of commodity prices on CPI inflation than GDP inflation and a larger rise in CPI volatility in 1991 that may reflect a shorter sample for estimation than is available with the GDP price index.

- As expected, the full sample (smoothed) SV estimates are modestly smoother than the quasi-real time (QRT) estimates. One dimension of this smoothness is that the QRT estimates tend to respond to recessions with a little delay; around recessions, the full sample estimates rise sooner than do the QRT estimates. In addition, in the case of CPI inflation, the late 2000's rise in volatility is larger in quasi-real time than in the full sample estimates.

5.2 Out-of-Sample Forecasts

As noted above, to assess forecast accuracy, we consider both interval forecasts and density accuracy as measured by the CRPS. We begin with the interval forecasts. Figures 8-12 report the forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach applied to a 60 observation rolling window of forecast errors and the other (in red) obtained from our stochastic volatility model of η_t. (We focus on forecast errors for simplicity; instead reporting the point forecasts and confidence bands around the forecasts would yield the same findings.) Again, for readability, we omit from the charts the estimates for the 3-step ahead horizon. These charts provide a read on time variation in the width of confidence intervals and the accuracy of the two approaches. Table 1 quantifies empirical coverage rates. In the discussion below, we focus on 68 percent coverage rates, because there are far fewer observations available for evaluating accuracy further out in the tails of the distributions.

The charts of the time paths of 68 percent confidence intervals display the following broad patterns.

- Both types of estimates (constant variances with rolling windows and our SV-based estimates) display considerable time variation in the width of the intervals. For GDP growth, unemployment, and GDP inflation (for which the evaluation sample dates back to 1984), the width of the constant variance estimates progressively narrows over the first half of the sample, reflecting the increasing influence of the Great Moderation on the rolling window variance estimates. In contrast, for CPI inflation, for which the sample is also shorter, the constant variance bands tend to widen as the sample moves forward.

- Consistent with the SV estimates discussed above, the width of the confidence bands based on our SV model-based approach varies more than does the width of intervals based on constant variances. For GDP growth, unemployment, and GDP inflation, the SV model-based intervals narrow sharply in the first part of the sample (more so than the constant variance estimates) and then widen significantly (again, more so than the constant variance estimates) with the Crisis and, in the case of GDP growth and the T-bill rate, the recession of 2001. For most of the sample, the interval widths are narrower with the SV approach than the constant variance approach; however,

this pattern does not so generally apply to CPI inflation.

- Across horizons, the contours of the confidence intervals (for a given approach) are very similar. With the SV model-based estimates, the similarities across horizons are particularly strong for horizons 1 through 4 (omitting the nowcast horizon).[24] Although the intervals display some differences in scales, they move together across horizons. In the model estimates, this comovement is reflected in estimates of the volatility innovation variance matrix Φ, which allows and captures some strong correlation in volatility innovations across horizons.[25] More broadly, with these variance estimates reflecting forecast uncertainty, as uncertainty varies over time, that uncertainty likely affects all forecast horizons, in a way captured by these SV estimates.

The coverage rates reported in Table 1 for SPF forecasts quantify the accuracy of the 68 percent confidence intervals shown in Figures 8-12.[26] These show the intervals based on our stochastic volatility model to be consistently more accurate than the intervals based on the constant variance approach applied to forecast errors. Although we cannot claim that the SV-based approach yields correct coverage in all cases, it does so in the large majority cases; the gap between the empirical and nominal rate is significant only in the case of TBILL forecasts are short horizons. Moreover, the SV-based approach typically improves on the alternative approach, which in most cases yields coverage rates above 68 percent, reflecting bands that are too wide. For example, for GDP growth, the SV-based coverage rates range (across horizons) from 69.5 percent to 73.6 percent, with no departures from 68 percent large enough to be statistically significant, whereas the constant variance-based rates range from 76.5 percent to 79.7 percent, with all five departures from 68 percent large enough to be statistically significant. For the T-bill rate, the SV-based rates are much lower than the constant-variance-based rates at forecast horizons of 2 quarters or more — e.g., at the 2-step horizon, 71.6 percent with SV versus 83.95 percent for the constant variance baseline. For the inflation measures considered, results for the GDP price index are comparable to

---

[24]Note that, for the unemployment and T-bill rates, the interval widths for the nowcast are narrower than those at longer horizons probably due to data timing, with forecasters often (unemployment) or always (T-bill rate) having available one month of data on the quarter.

[25]In unreported results, we have used the estimates of Φ at each forecast origin to construct the correlation matrix of the innovations to the volatility processes of the model. These estimates are fairly stable over time, with a largest eigen value between 2 and 3 (depending on the variable), the second largest eigen value equal to about 1, and the remaining eigen values progressively smaller, but none close to 0.

[26]Specifically, we use a nominal coverage rate of 68.27 percent, the percentile associated with a normal distribution and an error equal to one standard deviation.

those for real GDP. But for CPI inflation, the coverage rates obtained with our SV model are similar to those obtained with the constant variance benchmark approach.

To provide a broader assessment of density forecast accuracy, Table 2 reports the average CRPS. To simplify comparison, the table reports the level of the CRPS obtained with the constant variance approach and the percentage improvement in the CRPS of the SV-based forecasts relative to the constant variance-based forecasts. With SPF forecasts, for all variables, our SV model consistently offers density accuracy gains over the constant variance specification. The gains are largest for the T-bill rate, ranging from 5 to 14 percent. For GDP growth, the gains are still healthy, ranging from 4 to 11 percent. For the unemployment rate and the inflation measures, the gains are smaller, but consistently positive, ranging from 1.9 to 5.5 percent. As noted above, although some studies have found modestly larger density gains associated with SV, these studies typically commingle benefits to point forecasts with benefits to the variance aspect of the density forecasts. In our case, the point forecasts are the same across the approaches, so any gains in density accuracy come entirely from variance-related aspects of the forecast distribution.

5.2.1 Robustness to Rolling Window Choice

We have also examined the performance of SV against the constant variance approach with the rolling window underlying the constant variance specification either shorter or longer than the 60 observation setting of our baseline results. One alternative is to lengthen the rolling window to 80 observations, in line with the roughly 20 year sample underlying the historical forecast RMSEs reported in the Federal Reserve's SEP (Table 2).[27] Note, however, that with this setting, the comparison to the baseline isn't entirely a clean one, because the longer rolling window shortens (pulling the start point forward) the forecast evaluation sample by 20 observations.

Lengthening the rolling window underlying the benchmark constant variance approach does not alter the picture we painted above: the constant variance approach commonly yields coverage rates in excess of the nominal rate of 68 percent (see Appendix Table 1). In many cases, the coverage rates are higher with the 80 observation window than the 60 observation window. With the exception of CPI inflation, the empirical coverage rates are almost all significantly above 68 percent. Lengthening the rolling window underlying the benchmark constant variance approach does not alter the broader picture of density forecast

---

[27]For details, see http://www.federalreserve.gov/foia/files/20140409-historical-forecast-errors.pdf.

accuracy characterized above for the CRPS. In some cases (see Appendix Table 2), the gains of SV over the constant variance approach are a little smaller with the rolling window of 80 observations than the baseline window of 60 observations, but the changes do not appear material, and it remains the case that our SV specification offers consistent gains to CRPS accuracy over the constant variance approach.

The other alternative we consider shortens the rolling window to 40 observations. With this window setting, we could lengthen the evaluation sample, but to facilitate comparisons to the baseline case, we leave the evaluation sample unchanged from the baseline. In the estimates, shortening the rolling window to 40 observations does not materially change the picture provided by the baseline results, although in forecast coverage, it slightly reduces the advantage of our SV-based model. As detailed in the appendix (Appendix Table 3), coverage rates fall some for GDP growth and GDP inflation, making rates for the constant variance benchmark broadly comparable to those for the SV model. However, even with the shorter observation window for the constant variance approach, the SV model yields better coverage for the unemployment and T-bill rates. Density accuracy as measured by the CRPS is at most modestly affected (see Appendix Table 4) by shortening the rolling window from 60 to 40 observations, such that our SV model-based approach maintains the same consistent advantage described above.

5.2.2 Out-of-Sample Results for VAR-SV specification

In the interest of brevity, in examining the efficacy of extending our baseline SV model to the VAR-SV specification, we present the out-of-sample results and omit figures with the full-sample VAR-SV estimates of volatility. The full-sample estimates for the VAR-SV model are qualitatively similar to the baseline SV estimates. Tables 3 and 4 provide 68 percent coverage rates and CRPS values for the VAR-SV model, with comparison to the baseline constant forecast error variance approach (repeating these results from Tables 1 and 2 for convenience).

The coverage rates reported in Table 3 for SPF forecasts show the intervals based on the VAR-SV model to be modestly more accurate than the intervals based on the constant variance approach applied to forecast errors. In broad terms, the advantages of the VAR-SV model over the benchmark constant variance case can be seen in the number of asterisks, with fewer statistically significant departures from correct coverage. As examples, the VAR-SV model yields coverage rates much lower than the benchmark for the unemployment and

T-bill rates. However, in most cases, the advantages of the VAR-SV model are smaller than those of the baseline SV model. In most cases, coverage rates are higher with the VAR-SV model than the baseline SV model. This is associated with less accurate coverage in most cases, except for the T-bill rate.

For broader density forecast accuracy, the CRPS averages provided in Table 4 for SPF forecasts show the VAR-SV specification to be useful for some variables and not others. For GDP growth, the unemployment rate, and the T-bill rate, the VAR-SV model yields density forecasts more accurate than those obtained with the benchmark constant variance approach, with gains up to 6 percent for growth, up to 9 percent for unemployment, and up to 22 percent for the T-bill rate. For the inflation variables, the VAR-SV model yields density forecasts modestly less accurate than the benchmark. When compared to the baseline SV model, the extension provided by the VAR-SV model is somewhat helpful for unemployment and T-bill forecasts (boosting the CRPS noticeably) and somewhat harmful for the other variables.

On balance, this evidence suggests that extending our baseline SV model to depart from its MDS assumption has a mixed payoff. It helps along some, but not all, dimensions.

5.2.3 Out-of-Sample Results for Greenbook

In the interest of brevity, in examining the robustness of our results to the use of Greenbook rather than SPF forecasts, we present the out-of-sample results and omit figures with the full-sample $\eta_{\mathbf{t}}$ and SV estimates. The full-sample estimates with Greenbook are qualitatively similar to those for SPF.

Figures 13-16 report the Greenbook forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach applied to forecast errors and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model. The lower portion of Table 1 quantifies empirical coverage rates at the nominal rate of 68 percent.

In broad terms, along most dimensions, the pattern of interval forecast results for Greenbook are similar to those for SPF (in reviewing the charts, recall that the Greenbook sample ends almost six years earlier). First, as in the SPF results, both types of volatility estimates (constant variances with rolling windows and our SV-based estimates) display considerable time variation in the width of the intervals. However, in this dimension, the Greenbook results appear somewhat different is that, up to the mid-2000s, the bands around CPI in-

flation are fairly stable in width, whereas the SPF-based bands become gradually wider. Second, the width of the confidence bands based on our SV approach varies more than does the width of intervals based on constant variances. For example, for most variables, the bands widen substantially with the Great Recession and with earlier recessions. Third, across horizons, the contours of the confidence intervals (for a given approach) are very similar.

The coverage rates for Greenbook forecasts reported in Table 1 quantify the accuracy of the 68 percent confidence intervals shown in Figures 13-16. On balance, the intervals based on our stochastic volatility model perform comparably to those based on the constant variance approach. For CPI inflation, coverage rates are moderately better at longer horizons with stochastic volatility than in the benchmark. For the unemployment rate, coverage rates also tend to be somewhat closer to the nominal size with the stochastic volatility model than the constant variance approach. But for GDP growth and inflation, coverage rates are quite similar across the two approaches The CRPS averages for Greenbook forecasts given in the lower part of Table 2 show that, in most cases, our SV model consistently offers some density accuracy gains over the constant variance specification. In broad terms, the gains are comparable to those observed with SPF forecasts, but a little smaller in most cases. For example, for GDP growth, the gains range from about 4 to 11 percent with SPF forecasts and 5 to 7 percent with Greenbook forecasts. For the extension to the VAR-SV models, the patterns in the Greenbook forecasts are similar to those described above for the SPF forecasts. In 68 percent coverage, extending the SV model to include VAR dynamics does not seem to help much. But in broader density accuracy as captured by the CRPS, the extension to include VAR dynamics reduces accuracy for inflation variables but mostly preserves or extends the gains from SV for other variables.

On balance, our main results obtained with SPF forecasts are corroborated by estimates with Greenbook forecasts, although perhaps the efficacy of stochastic volatility is modestly less compelling with Greenbook than with SPF forecasts.

6 Model extensions

Along several dimensions, the model and data could be extended to include additional features. In the interest of brevity, we don't pursue these extensions, but we briefly describe them, leaving them as subjects for future research.

First, the multivariate stochastic volatility model could be extended to allow fat tails in the conditional errors $\boldsymbol{\epsilon_t}$, drawing on the specification of Jacquier, Polson, and Rossi (2004) or an outlier-filtering approach of Stock and Watson (2016). Some macroeconomic studies have used fat-tailed SV specifications with time series or structural models, with varying success (e.g., Chiu, Mumtaz, Pinter 2015; Clark and Ravazzolo 2015; Curdia, Del Negro, Greenwald 2015). Stock and Watson (2016) find a related, mixture of normals approach to filtering inflation outliers to be helpful. We actually examined extending our model to include the Stock-Watson mixture. In our setting with SPF forecasts, this model extension helped to reduce the influence of some outliers on our stochastic volatility estimates but to have little effect on our baseline forecast results.

Second, the model could be extended to make use of forecasts from multiple sources. In Reifschneider and Tulip (2007, 2017) and the Federal Reserve's SEP, forecast accuracy is estimated by averaging the root mean square errors of a range of forecasts. In our framework, multiple forecasts could be exploited by treating each forecast source as a different measurement on a common volatility process. That is, the data vector $\boldsymbol{\eta_t}$ could be expanded to include multiple measurements of the nowcast error and each of the expectational updates, driven by a common set of the $H + 1$ volatility processes and conditional errors.[28]

Third, the multivariate stochastic volatility model could be extended to include all of the variables being forecast. We have proceeded on a variable-by-variable to keep the computations tractable. However, it is likely possible — although more complicated — to consider some variables jointly, with the model extended (as described above) to permit correlation across the nowcast errors and expectational updates of different variables.

Finally, there are some ways our model might in future research be extended to allow asymmetries in the forecast (error) distributions. As noted above, we have interpreted past periods of non-zero forecast errors as biases and removed them, allowing some slow time variation intended to be consistent with evidence in Croushore (2010), before estimation and forecast evaluation. However, some might interpret these past periods of non-zero errors as representing asymmetries in the forecast distribution that should be explicitly modeled. It could be possible to draw on the finance literature on stochastic volatility with asymmetries (see, e.g., Asai, McAleer, and Yu's (2006) review of multivariate SV models with asymmetries) to extend our macroeconomic model to allow asymmetries. Some might also believe

---

[28]The model could further be extended by easing up on the common factor restrictions by, for example, introducing some measurement noise to each forecast source.

it important to explicitly model asymmetries in variables such as the unemployment rate and T-bill rate, particularly given effective lower bounds on these variables and historical work such as Neftci (1984) and Montgomery, et al. (1998) on the evidence of asymmetries in the unemployment rate.

7   Conclusions

Motivated in part by central bank fan charts that use historical forecast errors to quantify the uncertainty around forecasts, this paper develops a multiple-horizon specification of stochastic volatility for forecast errors from sources such as SPF, Blue Chip, or the Fed's Greenbook, for the purpose of improving the accuracy of uncertainty estimates around the forecasts. Our approach can be used to form confidence bands around forecasts that allow for variation over time in the width of the confidence bands; the explicit modeling of time variation of volatility eliminates the need for somewhat arbitrary judgments of sample stability.

At each forecast origin, we have available the forecast error from the previous quarter and forecasts for the current quarter and the subsequent four quarters. To address the challenge of overlap in forecast errors across horizons, we formulate the model to make use of the current quarter (period t) nowcast error and the forecast updates for subsequent quarters (forecasts made in period t less forecasts made in period $t-1$). These observations reflect the same information as the set of forecast errors for all horizons. However, unlike the vector of forecast errors covering multi-step horizons, the vector containing the forecast updates is serially uncorrelated, under conventional assumptions that the forecasts represent a vector of conditional expectations. For this vector of observations, we specify a multiple-horizon stochastic volatility model that can be estimated with Bayesian MCMC methods. From the estimates, we are able to compute the time-varying conditional variance of forecast errors at each horizon of interest.

Estimates of the model with the full sample of forecasts display considerable historical variation in forecast error variances, at each forecast horizon. Consistent with evidence from the VAR and DSGE literatures, the forecast error variances shrink significantly with the Great Moderation and tend to rise — temporarily — with each recession, most sharply for the most recent Great Recession. To assess the performance of our approach in out-of-sample forecasting we assess forecast coverage rates and the accuracy of density forecasts as

29

measured by the continuous ranked probability score. We show that, by these measures, our proposed approach yields forecasts more accurate than those obtained using simple variances computed with rolling windows of forecast errors as in approaches such as Reifschneider and Tulip (2007, 2017).

References

Abel, Joshua, Robert Rich, Joseph Song, and Joseph Tracy (2016), "The Measurement and Behavior of Uncertainty: Evidence from the ECB Survey of Professional Forecasters," *Journal of Applied Econometrics* 31, 533-550.

Asai, Manabu, Michael McAleer, and Jun Yu (2006), "Multivariate Stochastic Rolatility: A Review," *Econometric Reviews* 25, 145-175.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2015), "The Measurement and Characteristics of Professional Forecasters' Uncertainty," *Journal of Applied Econometrics* 30, 1029-1046.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016), "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business and Economic Statistics* 34, 375-390.

Chiu, Ching-Wai (Jeremy), Haroon Mumtaz, and Gabor Pinter (2015), "Forecasting with VAR models: Fat Tails and Stochastic Volatility," Bank of England Working Paper No. 528.

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

Clark, Todd E. (2011), "Real-time Density Forecasts from BVARs with Stochastic Volatility," *Journal of Business and Economic Statistics* 29, 327-341.

Clark, Todd E. and Francesco Ravazzolo (2015), "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics* 30, 551-575.

Clements, Michael P. (2014), "Forecast Uncertainty Ex Ante and Ex Post: U.S. Inflation and Output Growth," *Journal of Business and Economic Statistics* 32, 206-216.

Clements, Michael P. (2016), "Are Macroeconomic Density Forecasts Informative?" Discussion Paper ICM-2016-02, Henley Business School, University of Reading.

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.

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.

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

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

Croushore, Dean (2010), "An Evaluation of Inflation Forecasts from Surveys Using Real-

Time Data," *BE Journal of Macroeconomics: Contributions* 10, Article 10.

Croushore, Dean, and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130.

Curdia, Vasco, Marco Del Negro, and Daniel Greenwald (2015), "Rare Shocks, Great Recessions," *Journal of Applied Econometrics* 29, 1031-1052.

D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic Forecasting and Structural Change," *Journal of Applied Econometrics* 28, 82-101.

D?Amico Stefania, and Athanasios Orphanides (2008), "Uncertainty and Disagreement in Economic Forecasting," Finance and Economics Discussion Series Working Paper 2008-56, Federal Reserve Board of Governors.

Del Negro, Marco, and Giorgio E. Primiceri (2015), "Time Varying Structural Vector Autoregressions and Monetary Policy: a Corrigendum," *Review of Economic Studies* 82, 1342-1345.

Diebold, Francis X., Frank Schorfheide, and Minchul Shin (2016), "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," *Journal of Econometrics*, forthcoming.

Durbin, J., and S.J. Koopman (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika* 89, 603-615.

Faust, Jon, and Jonathan H. Wright (2008), "Efficient Forecast Tests for Conditional Policy Forecasts," *Journal of Econometrics* 146, 293-303.

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.

Giordani, Paolo, and Paul Soderlind (2003), "Inflation Forecast Uncertainty," *European Economic Review* 47, 1037-1059.

Giordani, Paolo, and Mattias Villani (2010), "Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction," *International Journal of Forecasting* 26, 312-325.

Gneiting, Tilmann, and Adrian E. Raftery (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association* 102, 359-378.

Gneiting, Tilmann, and Roopesh Ranjan (2011), "Comparing Density Forecasts Using Threshold and Quantile Weighted Proper Scoring Rules," *Journal of Business and Economic Statistics* 29, 411-422.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (2004), Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors," *Journal of Econometrics* 122, 185-212.

Justiniano, Alejandro and Giorgio E. Primiceri (2008), "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review* 93, 604-641.

Kenny, Geoff, Thomas Kostka, and Federico Masera (2014), "How Informative are the

Subjective Density Forecasts of Macroeconomists?" *Journal of Forecasting* 33, 163-185.

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

Knuppel, Malte (2014), "Efficient Estimation of Forecast Uncertainty Based on Recent Forecast Errors," *International Journal of Forecasting* 30, 257-267.

Kozicki, Sharon, and Barak Hoffman (2004), "Rounding Error: A Distorting Influence on Index Data," *Journal of Money, Credit, and Banking* 36, 319-338.

Krueger, Fabian, Sebastian Lerch, Thordis L. Thorarinsdottir, and Tilmann Gneiting (2017), "Probabilistic Forecasting and Comparative Model Assessment Based on Markov Chain Monte Carlo Output," manuscript.

Lahiri, Kajal, and Xuguang Sheng (2010), "Measuring Forecast Uncertainty by Disagreement: the Missing Link," *Journal of Applied Econometrics* 25, 514-538.

Mertens, Elmar, and James M. Nason (2015), "Time-varying Stickiness in Professional Inflation Forecasts," manuscript.

Montgomery, Alan L., Victor Zarnowitz, Ruey S. Tsay, and George C. Tiao (1998), "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association* 93, 478-93.

Neftci, Salih N. (1994), "Are Economic Time Series Asymmetric over the Business Cycle?" *Journal of Political Economy* 92, 307-328.

Newey, Whitney K., and Kenneth D. West (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

Patton, Andrew J., and Allan Timmermann (2012), "Forecast Rationality Tests Based on Multi-Horizon Bounds," *Journal of Business and Economic Statistics* 30, 1-17.

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

Ravazzolo, Francesco and Shaun P. Vahey (2014), "Forecast Densities for Economic Aggregates from Disaggregate Ensembles," *Studies in Nonlinear Dynamics & Econometrics* 18, 367-381.

Reifschneider, David, and Peter Tulip (2007), "Gauging the Uncertainty of the Economic Outlook from Historical Forecasting Errors," FEDS working paper 2007-60, Federal Reserve Board of Governors.

Reifschneider, David, and Peter Tulip (2017), "Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve's Approach," FEDS working paper 2017-20, Federal Reserve Board of Governors.

Rossi, Barbara, and Tatevik Sekhposyan (2014), "Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set," *International Jour-*

nal of Forecasting 30, 662-682.

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.

Stock, James H., and Mark W. Watson (2016), "Core Inflation and Trend Inflation," *Review of Economics and Statistics* 98, 770-784.

Table 1: 68 percent forecast error coverage rates

Variable	SPF horizon					eval. begin
	0	1	2	3	4	
Panel A: SV						
RGDP-SPF	72.73	71.76	72.31	72.87	69.53	1984:Q1
UNRATE-SPF	74.63	75.19	69.70	66.41	63.85	1984:Q1
PGDP-SPF	73.48	72.52	73.85	71.32	71.09	1984:Q1
CPI-SPF	68.67	64.63	64.20	67.50	69.62	1996:Q4
TBILL-SPF	80.72**	80.49**	72.84	65.00	51.90	1996:Q4
RGDP-Greenbook	73.98	81.15***	76.86**	78.33**	82.35***	1981:Q1
Unrate-Greenbook	73.39	70.73	75.41	71.90	66.67	1981:Q1
PGDP-Greenbook	71.54	77.87**	77.69**	79.17**	78.15**	1981:Q1
CPI-Greenbook	69.12	67.16	66.67	67.69	68.75	1995:Q1
Panel B: FE-CONST						
RGDP-SPF	76.52**	78.63**	76.92*	78.29*	79.69**	1984:Q1
UNRATE-SPF	73.13	82.71***	87.12***	87.79***	86.92***	1984:Q1
PGDP-SPF	74.24	78.63***	77.69**	79.07**	79.69**	1984:Q1
CPI-SPF	71.08	63.41	67.90	66.25	70.89	1996:Q4
TBILL-SPF	79.52**	87.80***	83.95**	80.00	78.48	1996:Q4
RGDP-Greenbook	76.42**	80.33***	74.38	77.50	77.31	1981:Q1
Unrate-Greenbook	69.35	72.36	75.41	75.21	78.33*	1981:Q1
PGDP-Greenbook	75.61*	79.51***	76.03	78.33*	73.11	1981:Q1
CPI-Greenbook	64.71	65.67	60.61	55.38	57.81	1995:Q1

Note: The table reports the empirical coverage rates of 68 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2017:Q2, in case of the SPF, and through 2011:Q4, in case of the Greenbook (evaluated against realized data as far as available in both cases). The upper panel provides results based on our proposed multi-horizon SV model. The lower panel provides results based on a constant-variance model estimated over rolling windows with 60 quarterly observations. Statistically significant departures from correct nominal coverage are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 2: Density forecast accuracy as measured by CRPS

Variable	SPF horizon					eval. begin
	0	1	2	3	4	
RGDP-SPF						
(SV rel.)	3.04%	7.19%	7.55%	8.52%	6.33%	1984:Q1
(FE-CONST abs.)	0.83	1.02	1.11	1.17	1.19	
UNRATE-SPF						
(SV rel.)	0.91%	1.75%	2.48%	2.51%	1.56%	1984:Q1
(FE-CONST abs.)	0.08	0.17	0.25	0.34	0.44	
PGDP-SPF						
(SV rel.)	0.58%	1.61%	2.37%	2.43%	3.26%	1984:Q1
(FE-CONST abs.)	0.50	0.56	0.60	0.63	0.68	
CPI-SPF						
(SV rel.)	1.08%	1.14%	1.53%	2.65%	2.12%	1996:Q4
(FE-CONST abs.)	0.67	1.06	1.11	1.11	1.11	
TBILL-SPF						
(SV rel.)	8.65%	12.09%	11.20%	8.07%	5.19%	1996:Q4
(FE-CONST abs.)	0.08	0.24	0.41	0.59	0.77	
RGDP-Greenbook						
(SV rel.)	2.41%	5.25%	5.59%	5.57%	2.73%	1981:Q1
(FE-CONST abs.)	0.89	1.18	1.33	1.40	1.35	
Unrate-Greenbook						
(SV rel.)	1.13%	1.47%	2.59%	2.16%	1.20%	1981:Q1
(FE-CONST abs.)	0.10	0.20	0.30	0.40	0.50	
PGDP-Greenbook						
(SV rel.)	1.36%	0.48%	1.86%	2.35%	3.26%	1981:Q1
(FE-CONST abs.)	0.54	0.60	0.61	0.65	0.69	
CPI-Greenbook						
(SV rel.)	3.60%	2.90%	4.93%	3.80%	3.90%	1995:Q1
(FE-CONST abs.)	0.55	1.11	1.23	1.25	1.23	

Note: The table reports CRPS results for out-of-sample density forecasts. The sample uses predictions made from the date given in the right-most column through 2017:Q2, in case of the SPF, and through 2011:Q4, in case of the Greenbook (evaluated against realized data as far as available in both cases). For each variable, the top row reports the relative CRPS calculated as the percentage decrease of the CRPS when using SV over constant variance (relative to the constant-variance CRPS); positive numbers indicate improvement of SV over the constant-variance case. The bottom row reports the CRPS for the constant-variance case, in which case the constant-variance model is estimated over rolling windows with 60 quarterly observations.

Table 3: 68 percent forecast error coverage rates, VAR-SV specification

| Variable	SPF horizon					eval. begin
	0	1	2	3	4	
Panel A: VAR-SV						
RGDP-SPF	74.24	76.34*	74.62	75.97	74.22	1984:Q1
UNRATE-SPF	72.39	76.69*	75.00	78.63**	73.08	1984:Q1
PGDP-SPF	73.48	77.10**	80.00***	78.29**	79.69***	1984:Q1
CPI-SPF	71.08	70.73	71.60	75.00	77.22**	1996:Q4
TBILL-SPF	71.08	80.49**	74.07	68.75	67.09	1996:Q4
RGDP-Greenbook	74.80*	83.61***	83.47***	82.50***	83.19***	1981:Q1
Unrate-Greenbook	70.97	73.98	77.05	77.69*	74.17	1981:Q1
PGDP-Greenbook	73.98	77.05**	81.82***	80.00***	82.35***	1981:Q1
CPI-Greenbook	70.59	70.15	69.70	72.31	73.44	1995:Q1
Panel B: FE-CONST						
RGDP-SPF	76.52**	78.63**	76.92*	78.29*	79.69**	1984:Q1
UNRATE-SPF	73.13	82.71***	87.12***	87.79***	86.92***	1984:Q1
PGDP-SPF	74.24	78.63***	77.69**	79.07**	79.69**	1984:Q1
CPI-SPF	71.08	63.41	67.90	66.25	70.89	1996:Q4
TBILL-SPF	79.52**	87.80***	83.95**	80.00	78.48	1996:Q4
RGDP-Greenbook	76.42**	80.33***	74.38	77.50	77.31	1981:Q1
Unrate-Greenbook	69.35	72.36	75.41	75.21	78.33*	1981:Q1
PGDP-Greenbook	75.61*	79.51***	76.03	78.33*	73.11	1981:Q1
CPI-Greenbook	64.71	65.67	60.61	55.38	57.81	1995:Q1

Note: The table reports the empirical coverage rates of 68 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2017:Q2, in case of the SPF, and through 2011:Q4, in case of the Greenbook (evaluated against realized data as far as available in both cases). The upper panel provides results based on our multi-horizon VAR-SV model. The lower panel provides results based on a constant-variance model estimated over rolling windows with 60 quarterly observations. Statistically significant departures from correct nominal coverage are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 4: Density forecast accuracy as measured by CRPS, VAR-SV specification

Variable			SPF horizon			eval. begin
	0	1	2	3	4	
RGDP-SPF						
(VAR-SV rel.)	1.86%	6.15%	4.46%	5.94%	3.02%	1984:Q1
(FE-CONST abs.)	0.83	1.02	1.11	1.17	1.19	
UNRATE-SPF						
(VAR-SV rel.)	9.43%	8.10%	7.70%	6.35%	3.80%	1984:Q1
(FE-CONST abs.)	0.08	0.17	0.25	0.34	0.44	
PGDP-SPF						
(VAR-SV rel.)	−2.01%	−2.88%	−2.62%	−3.99%	−5.14%	1984:Q1
(FE-CONST abs.)	0.50	0.56	0.60	0.63	0.68	
CPI-SPF						
(VAR-SV rel.)	−1.32%	−1.63%	−1.65%	−2.74%	−3.78%	1996:Q4
(FE-CONST abs.)	0.67	1.06	1.11	1.11	1.11	
TBILL-SPF						
(VAR-SV rel.)	19.90%	22.05%	22.53%	19.58%	17.06%	1996:Q4
(FE-CONST abs.)	0.08	0.24	0.41	0.59	0.77	
RGDP-Greenbook						
(VAR-SV rel.)	1.40%	5.36%	4.94%	4.75%	1.00%	1981:Q1
(FE-CONST abs.)	0.89	1.18	1.33	1.40	1.35	
Unrate-Greenbook						
(VAR-SV rel.)	5.61%	4.76%	4.61%	3.25%	1.40%	1981:Q1
(FE-CONST abs.)	0.10	0.20	0.30	0.40	0.50	
PGDP-Greenbook						
(VAR-SV rel.)	−0.96%	−5.81%	−10.37%	−14.25%	−17.22%	1981:Q1
(FE-CONST abs.)	0.54	0.60	0.61	0.65	0.69	
CPI-Greenbook						
(VAR-SV rel.)	2.37%	−2.41%	−0.28%	−1.89%	−2.87%	1995:Q1
(FE-CONST abs.)	0.55	1.11	1.23	1.25	1.23	

Note: The table reports CRPS results for out-of-sample density forecasts. The sample uses predictions made from the date given in the right-most column through 2017:Q2, in case of the SPF, and through 2011:Q4, in case of the Greenbook (evaluated against realized data as far as available in both cases). For each variable, the top row reports the relative CRPS calculated as the percentage decrease of the CRPS when using VAR-SV over constant variance (relative to the constant-variance CRPS); positive numbers indicate improvement of VAR-SV over the constant-variance case. The bottom row reports the CRPS for the constant-variance case, in which case the constant-variance model is estimated over rolling windows with 60 quarterly observations.

# Figure 1: Expectational Updates for Real GDP Growth
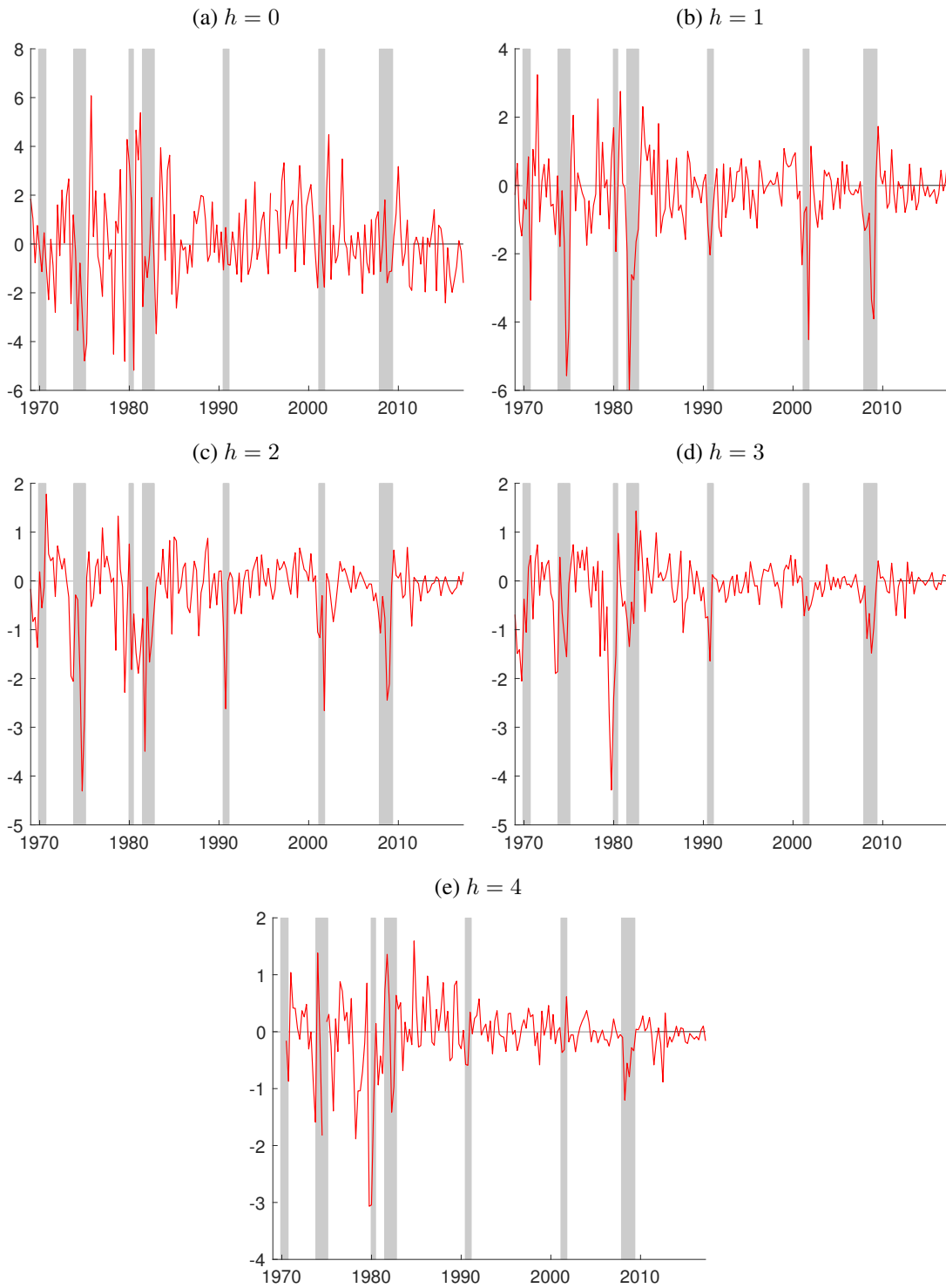
(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 3$

(e) $h = 4$

Note: The figure reports (in the red lines) the elements of the vector of expectational updates $\eta_{\mathbf{t}}$ used in model estimation. NBER recessions are indicated by gray bars.

Figure 2: Expectational Updates for the Unemployment Rate
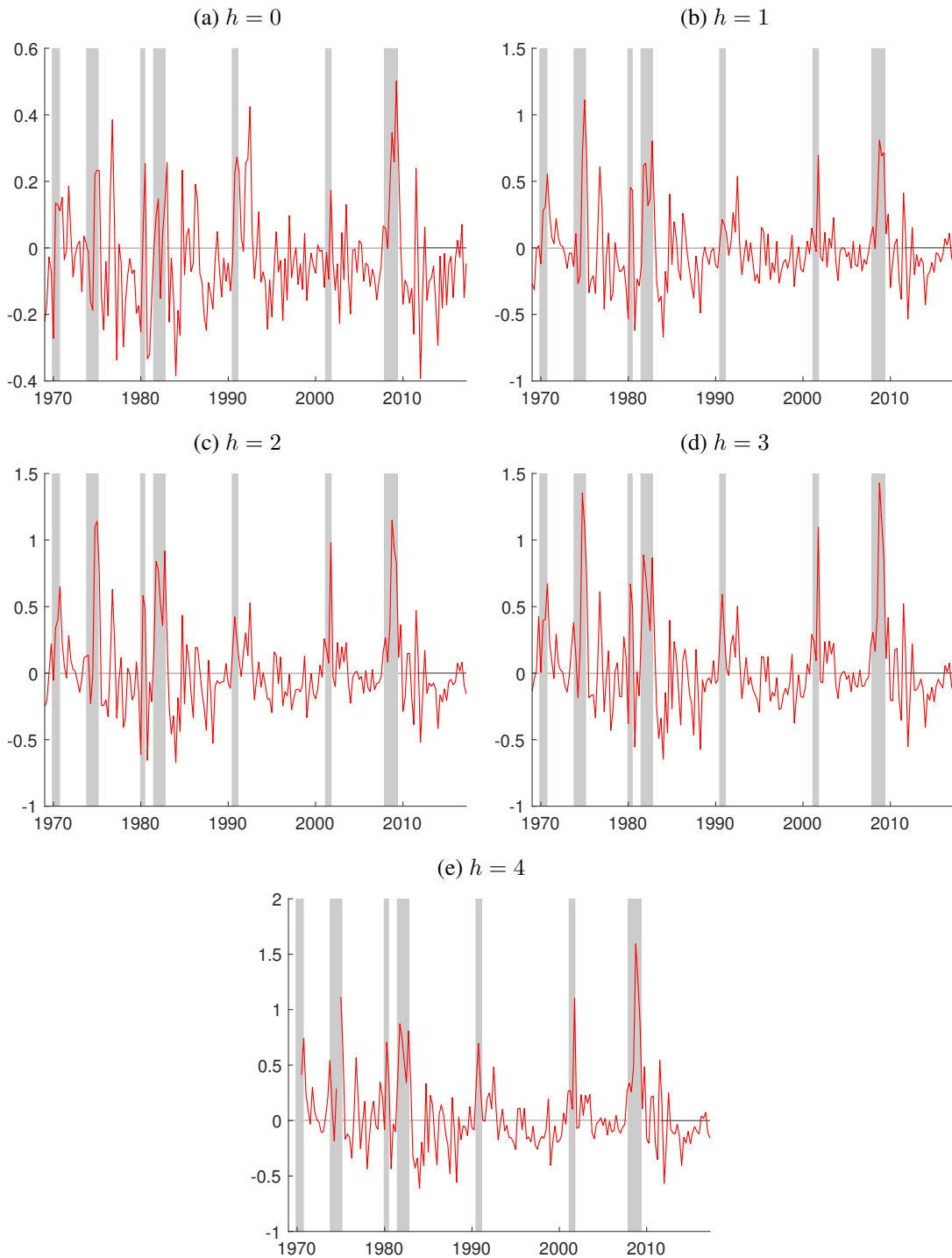
(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 3$

(e) $h = 4$

Note: The figure reports (in the red lines) the elements of the vector of expectational updates $\eta_{\mathbf{t}}$ used in model estimation. NBER recessions are indicated by gray bars.

Figure 3: Stochastic Volatility in Expectational Updates for Real GDP Growth
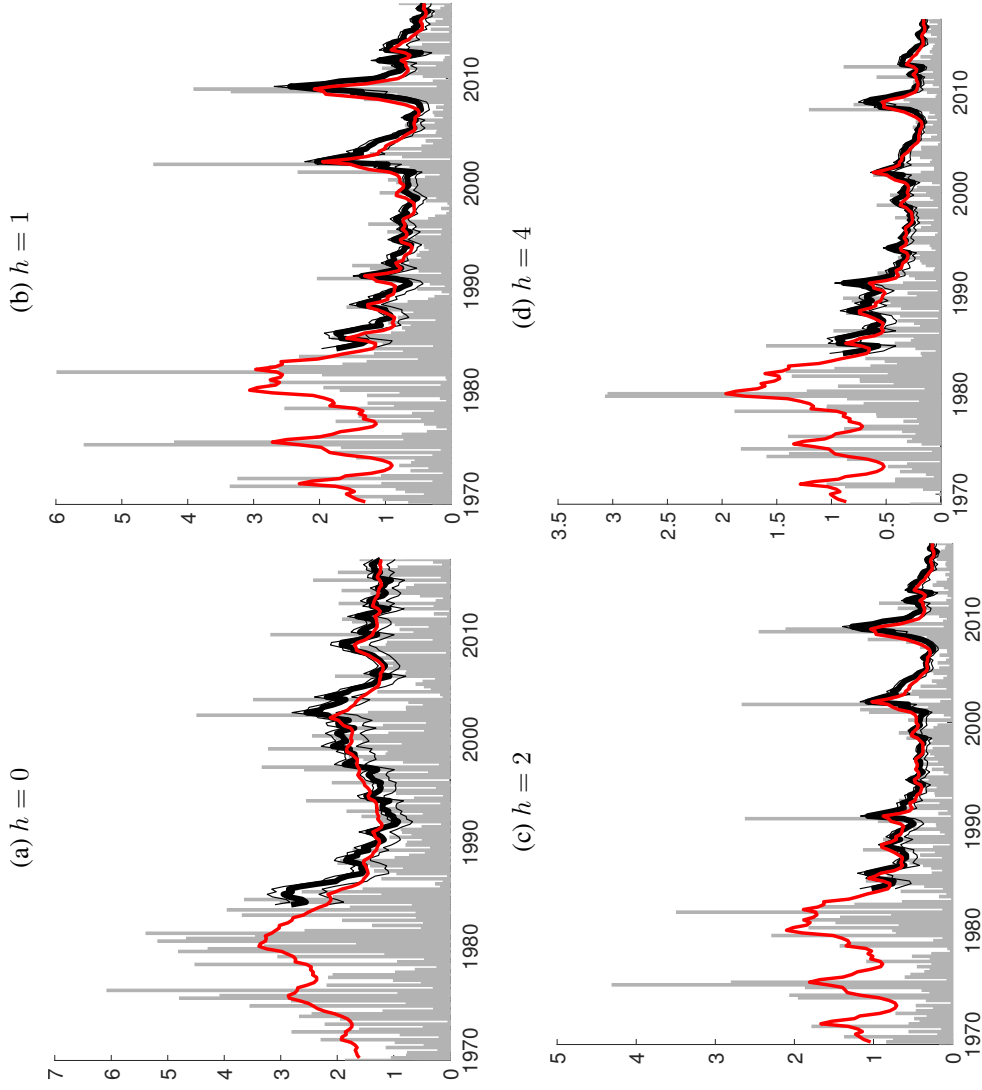
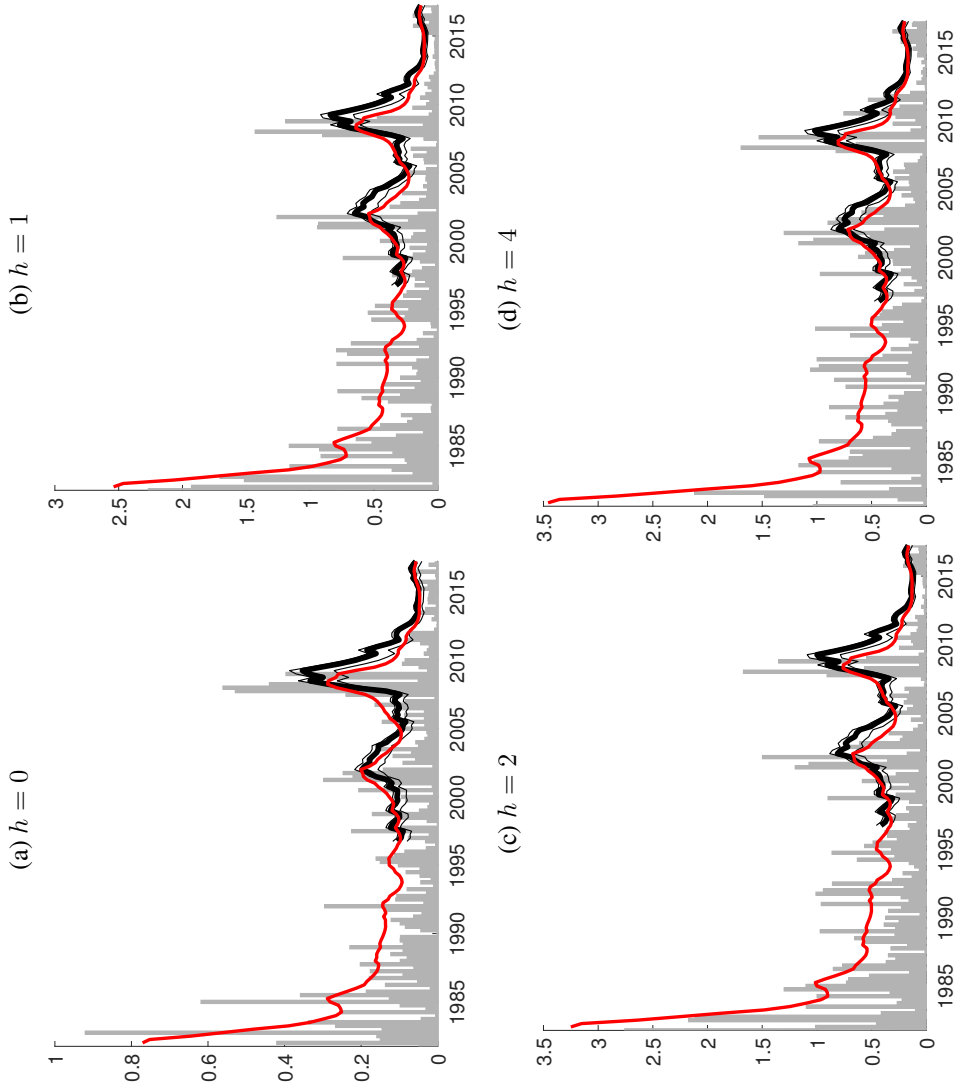(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating a volatility path at each forecast origin.

41

Figure 4: Stochastic Volatility in Expectational Updates for the Unemployment Rate
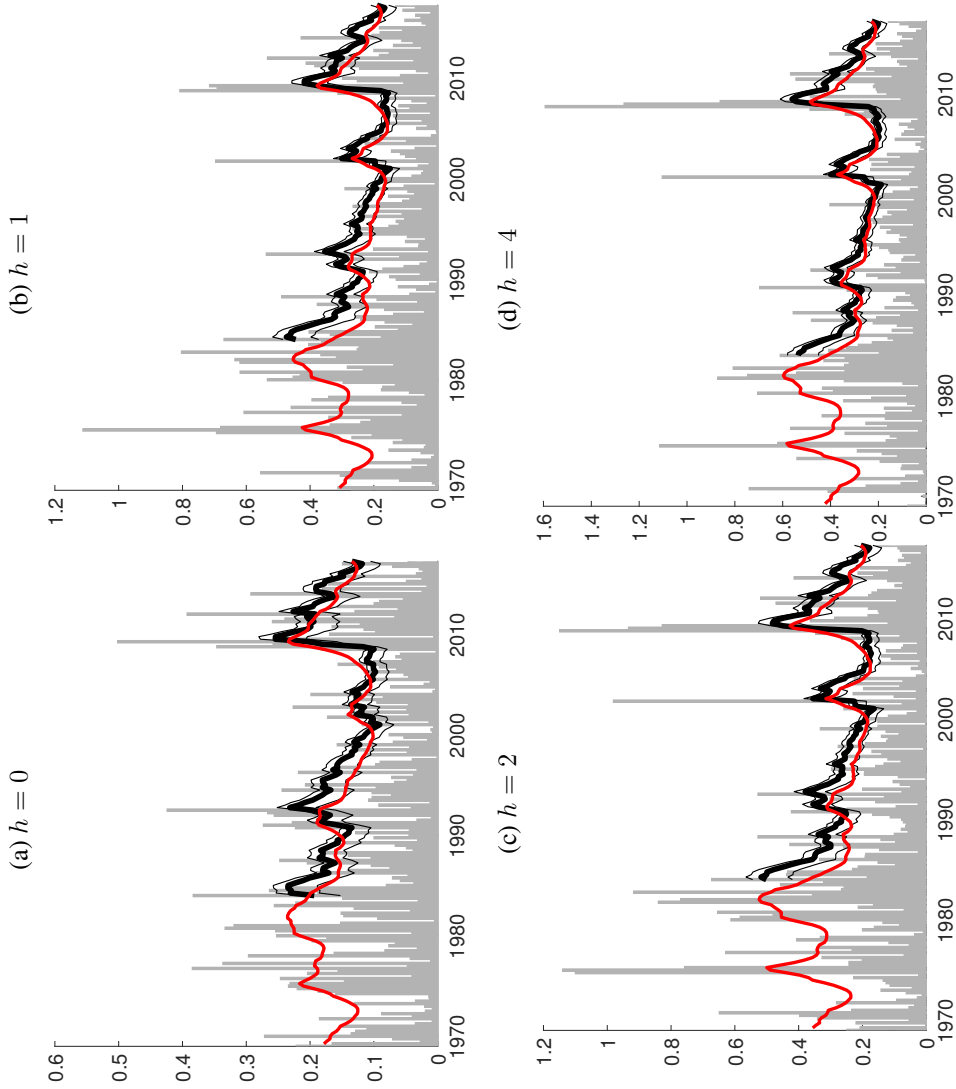
(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating a volatility path at each forecast origin.

42

Figure 5: Stochastic Volatility in Expectational Updates for GDP Deflator Inflation
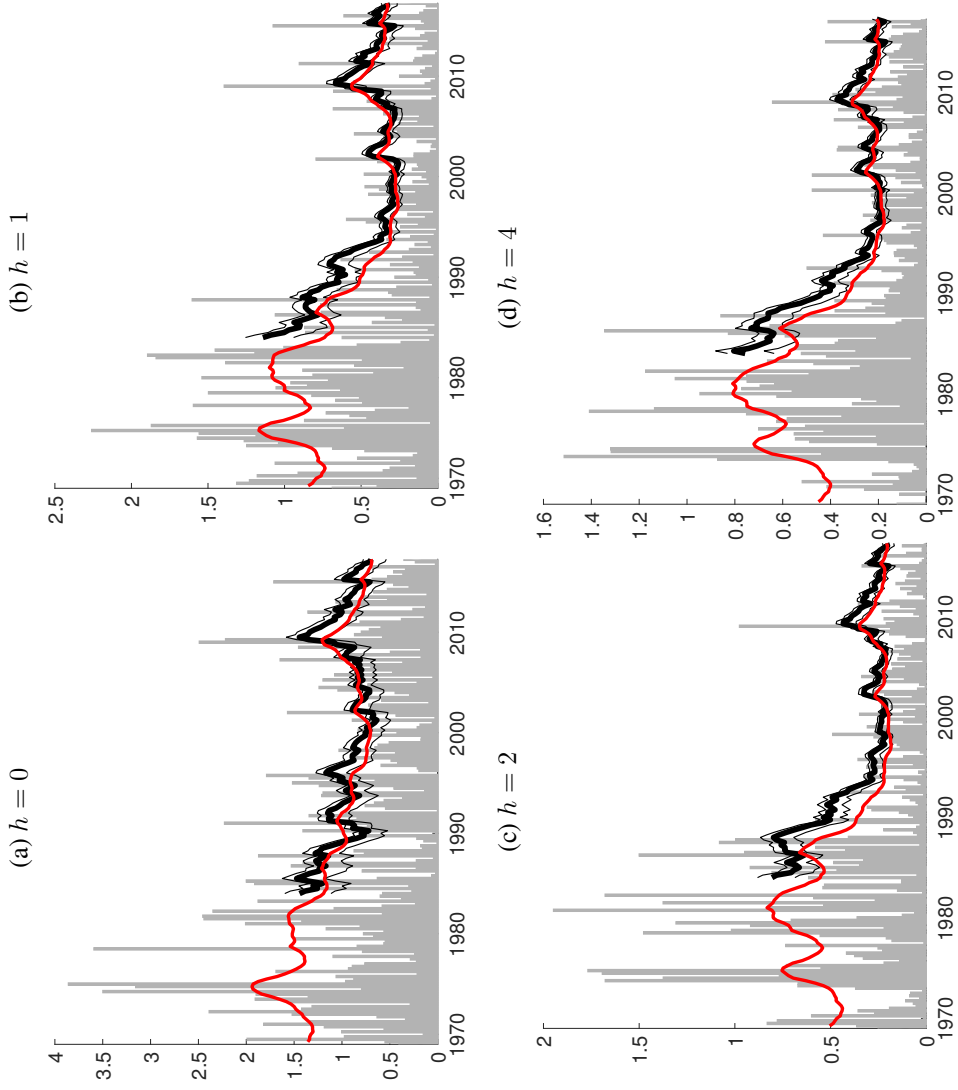
(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating a volatility path at each forecast origin.

43

Figure 6: Stochastic Volatility in Expectational Updates for CPI Inflation
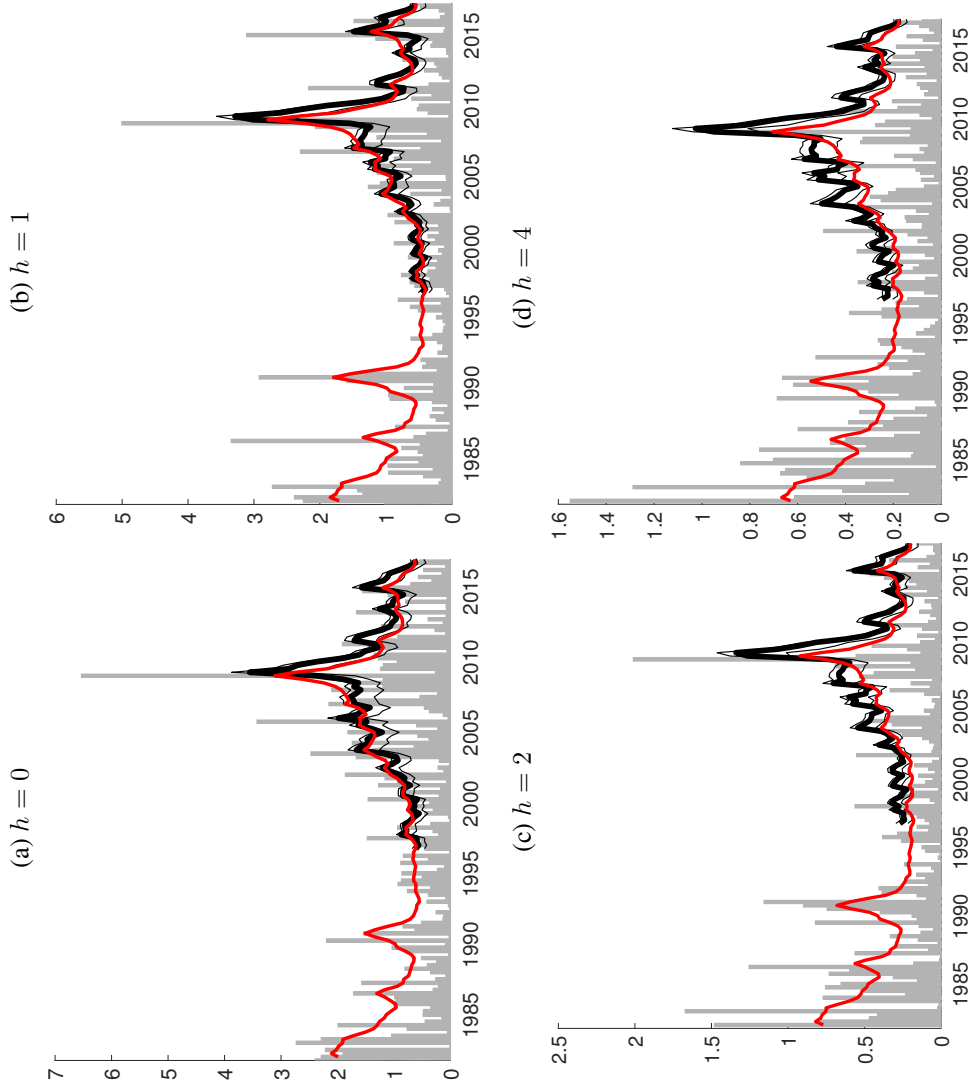
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating a volatility path at each forecast origin.

Figure 7: Stochastic Volatility in Expectational Updates for the Three-Month Tbill Rate

Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates, as well as quasi-real time estimates of stochastic volatility (black lines). The quasi-real time estimates are obtained by looping over time and estimating a volatility path at each forecast origin.

Figure 8: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for Real GDP Growth
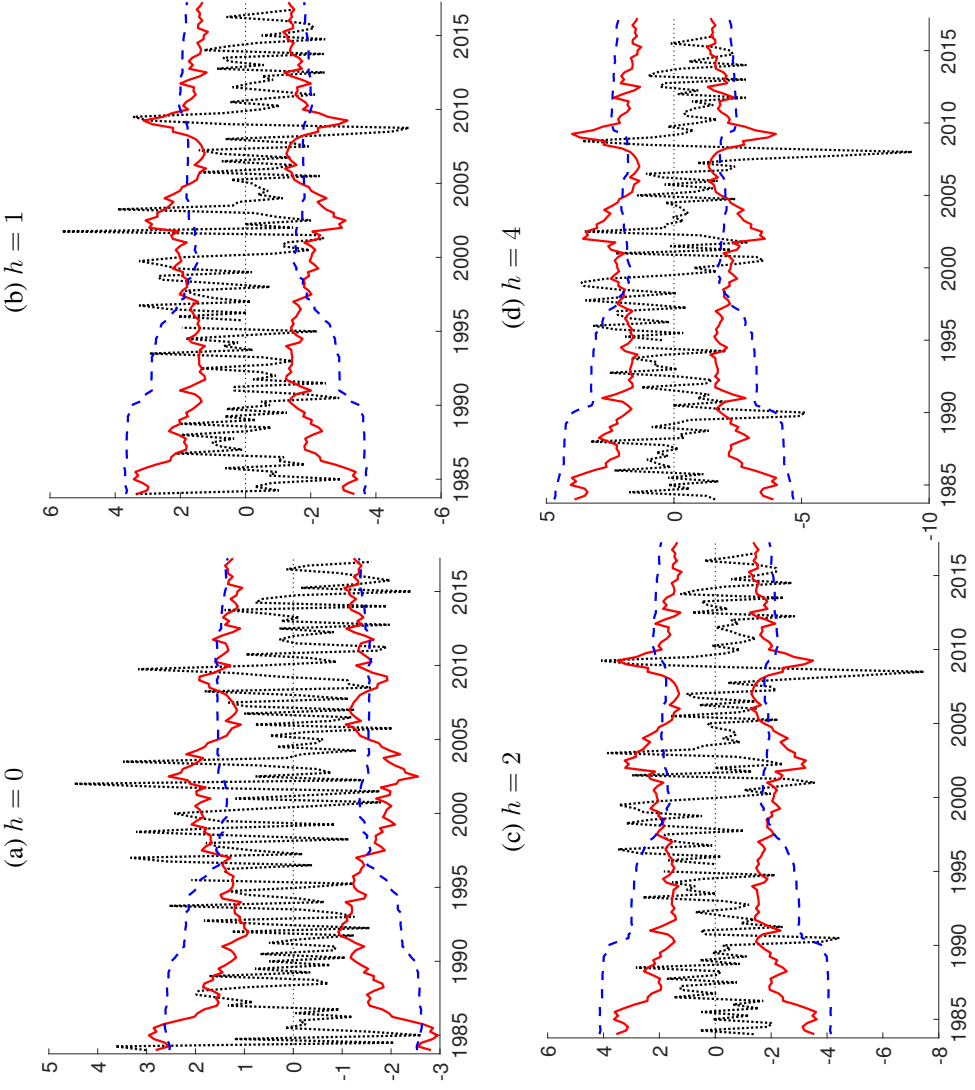
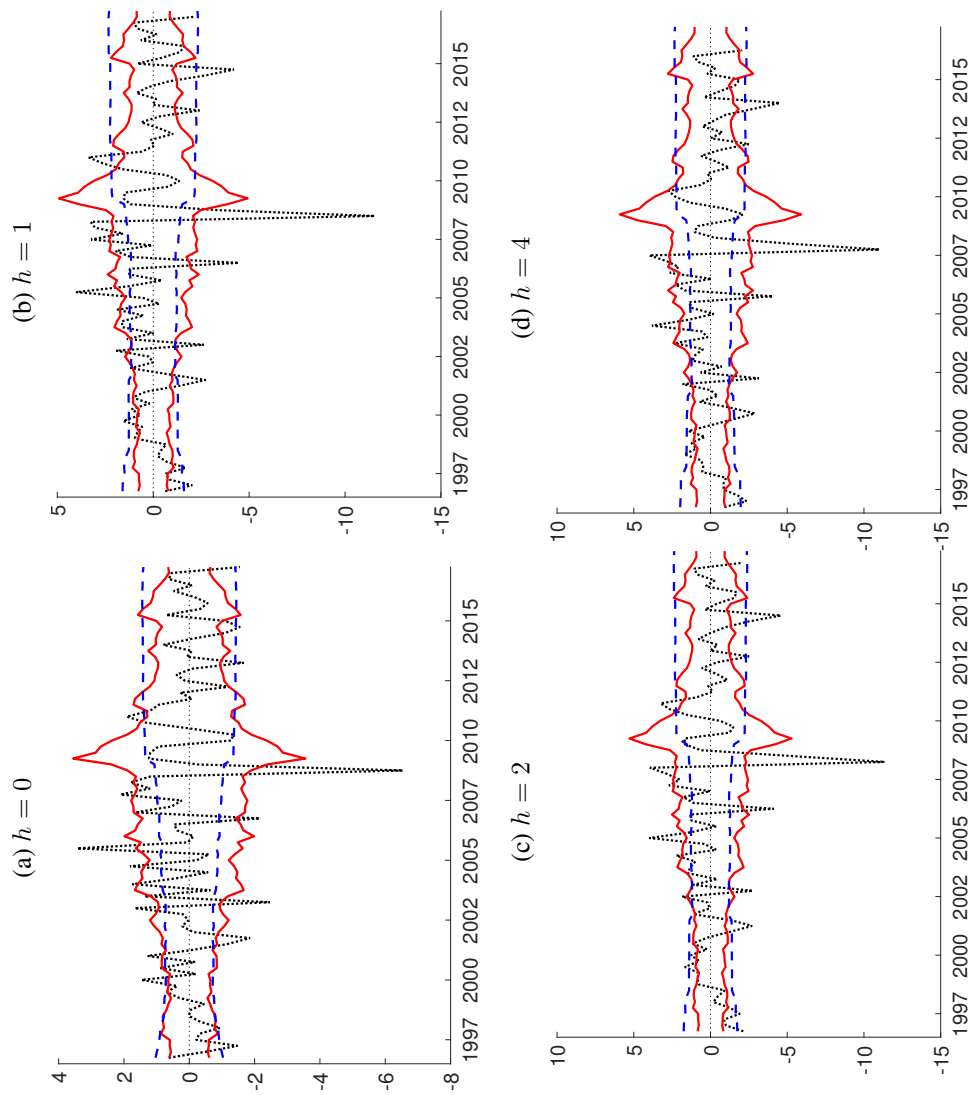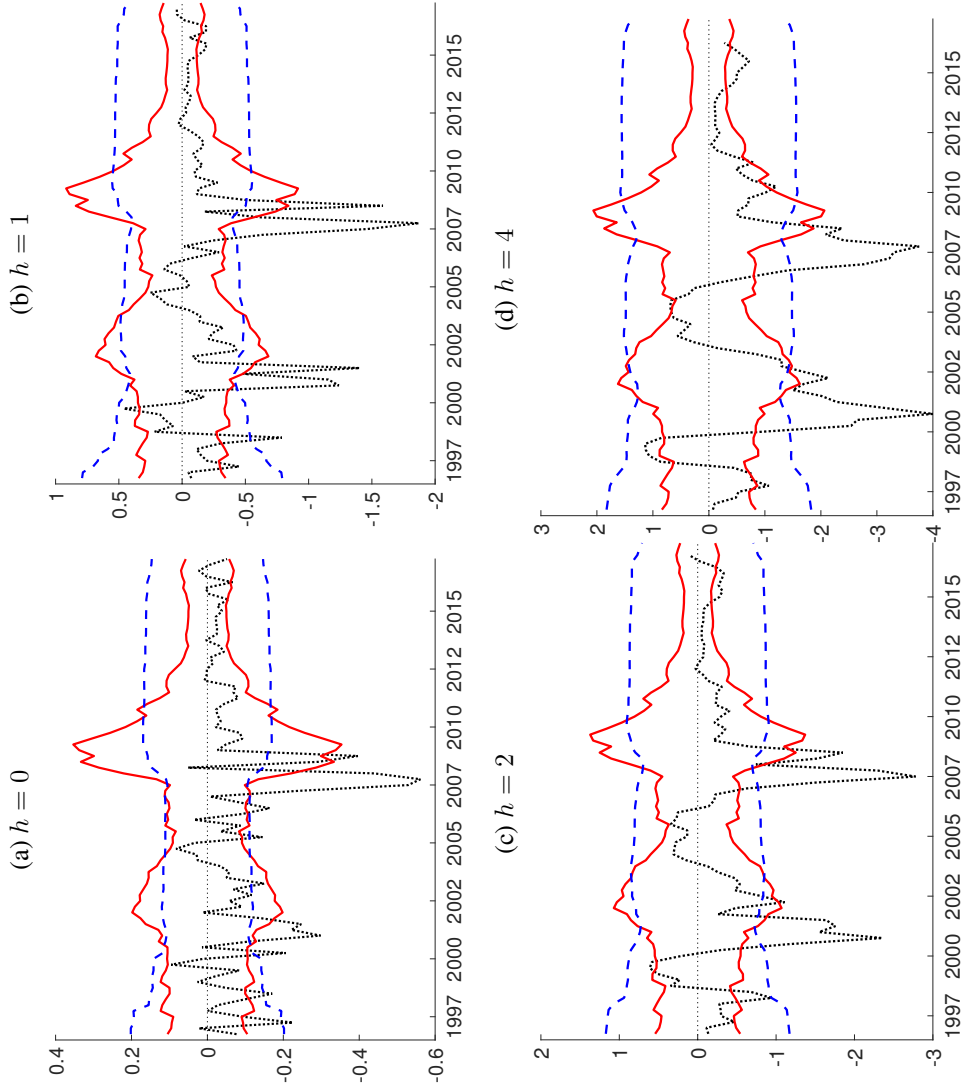(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

46

Figure 9: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for the Unemployment Rate
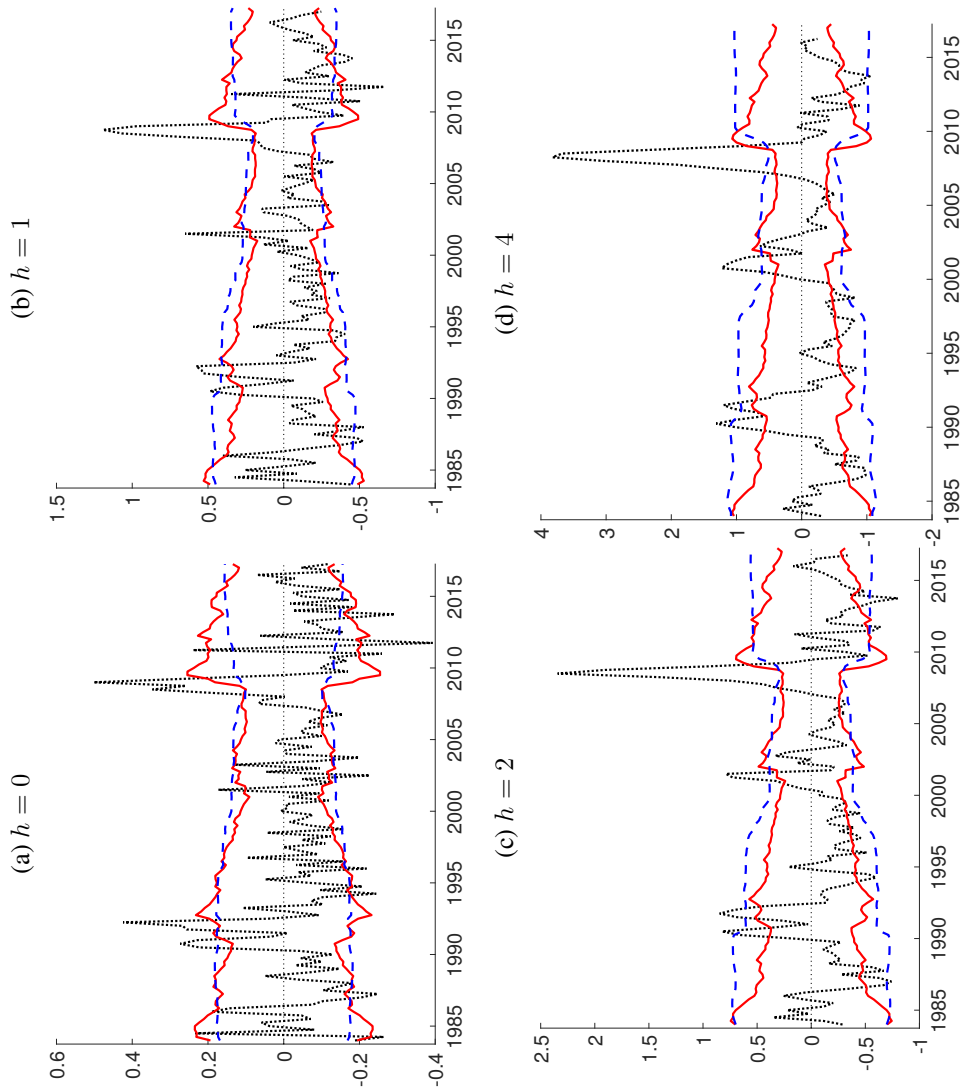
(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

47

Figure 10: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for GDP Deflator Inflation
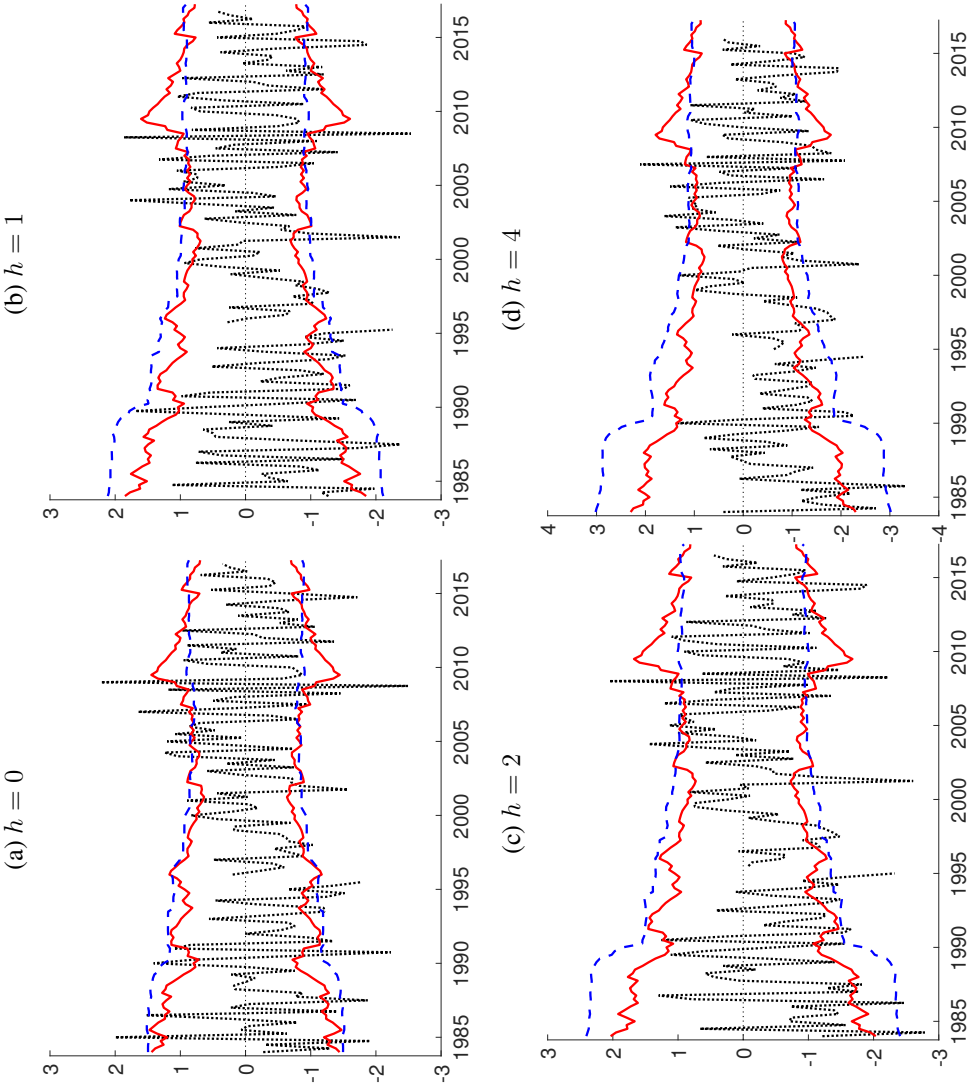
(a) $h = 0$

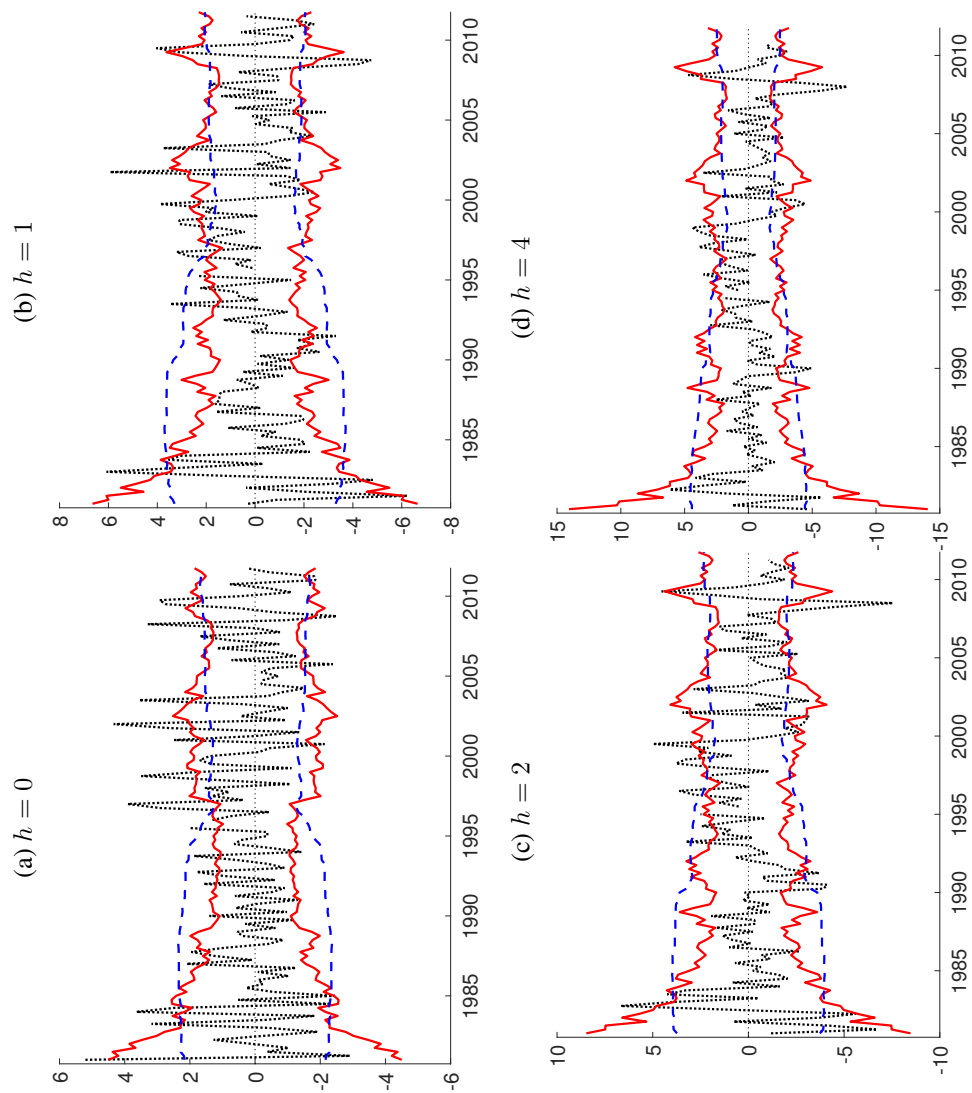(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 11: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for CPI Inflation
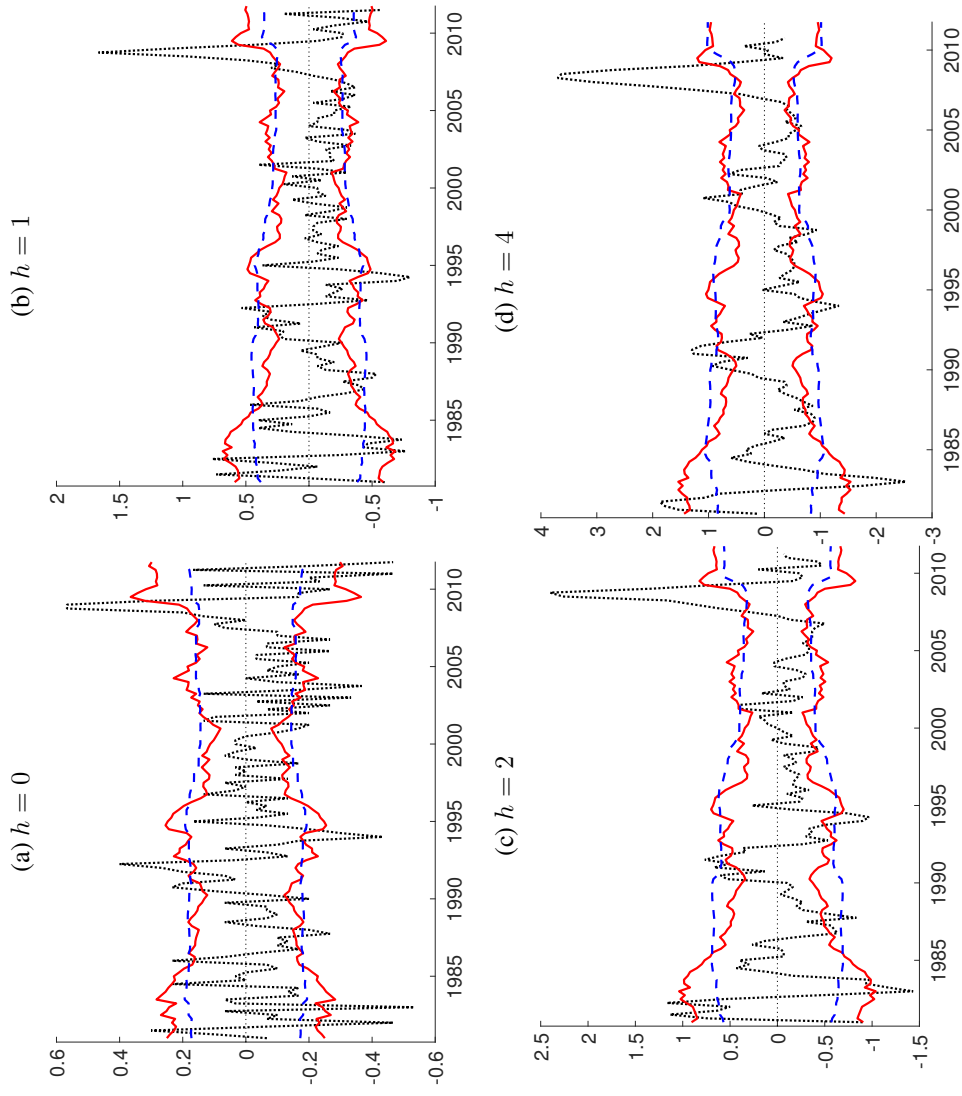
Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 12: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for the Three-Month Tbill Rate

(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 13: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for Real GDP Growth: Greenbook Forecasts

(a) $h = 0$
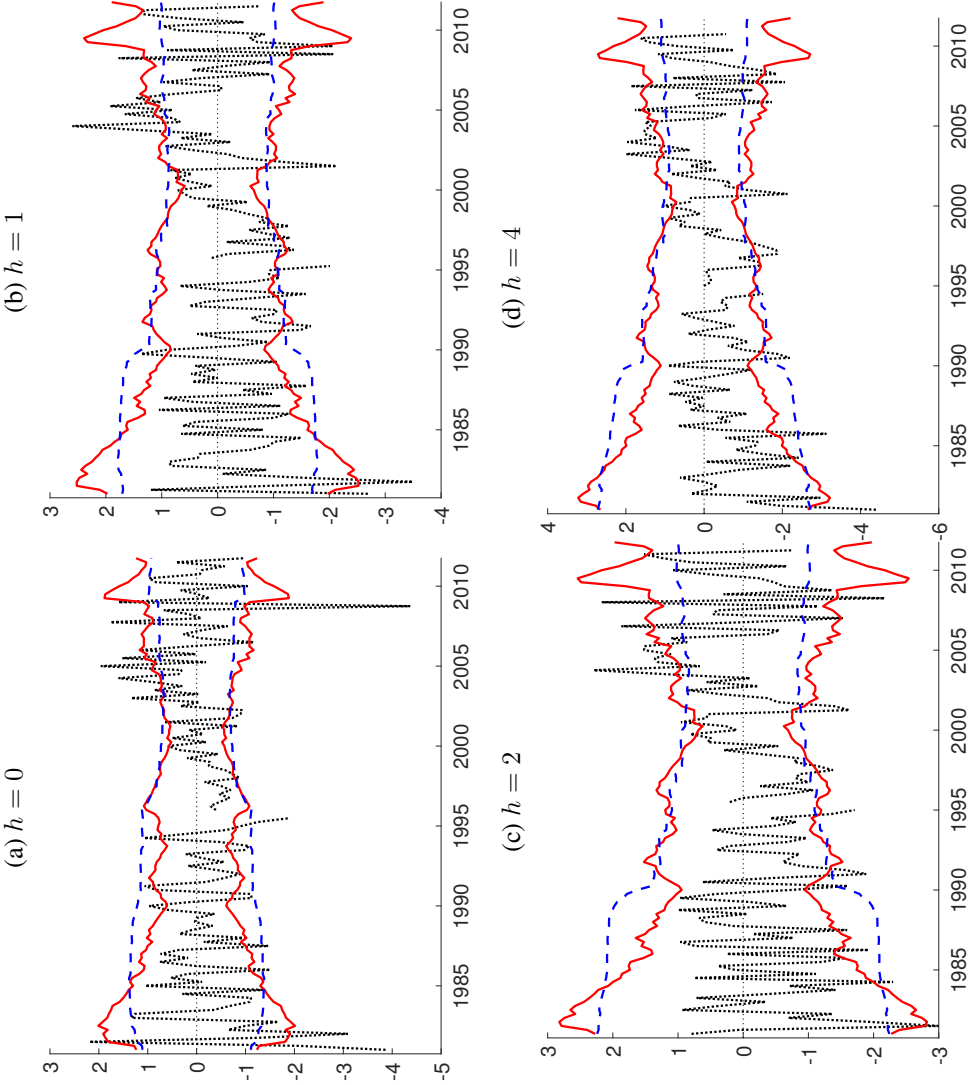
(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 14: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for the Unemployment Rate: Greenbook Forecasts
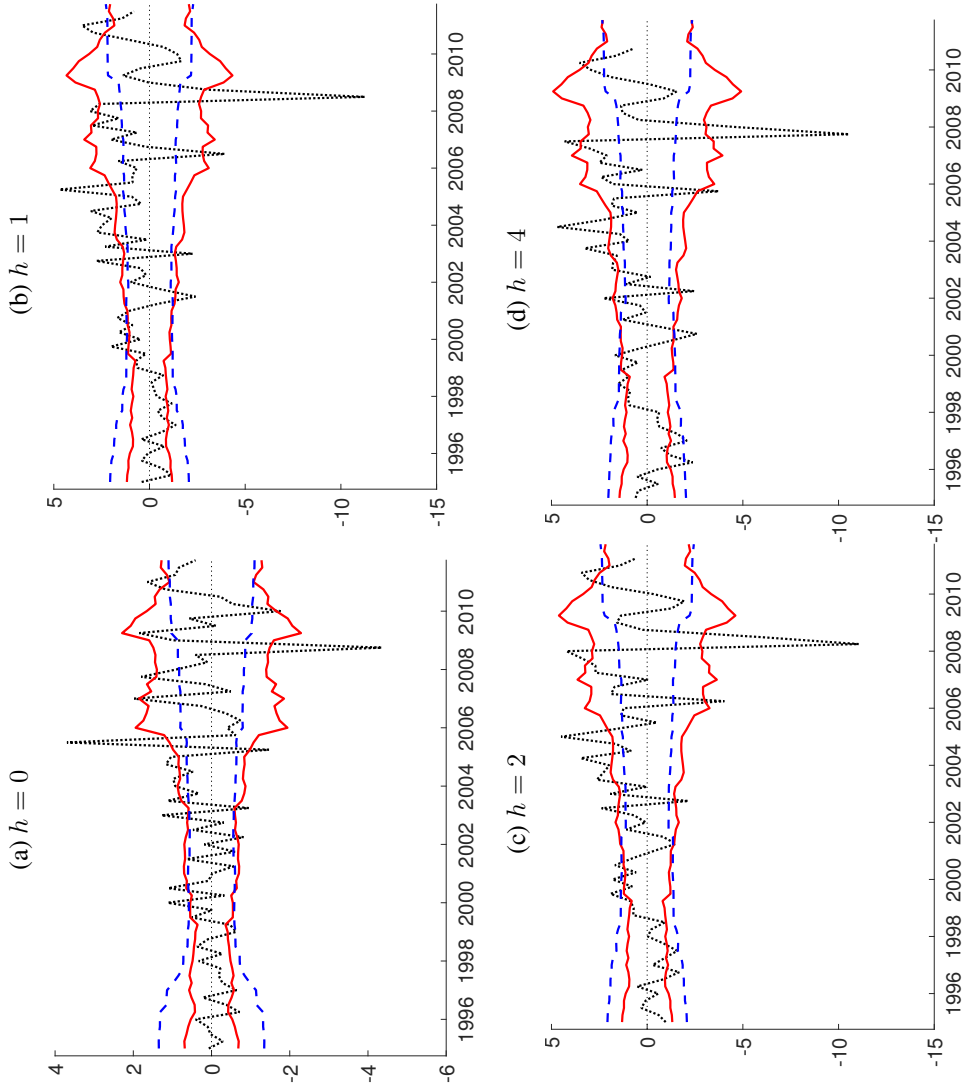

(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 15: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for GDP Deflator Inflation: Greenbook Forecasts

(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and the other (in red) obtained from our stochastic volatility model.

Figure 16: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for CPI Inflation: Greenbook Forecasts

(a) $h = 0$

(b) $h = 1$

(c) $h = 2$

(d) $h = 4$

Note: Each figure reports forecast errors for each variable along with 68 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.