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: March 14, 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.

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, Federal Reserve Bank of St. Louis, or the Bank for International Settlements.

1 Introduction

A number of central banks use the accuracy of historical forecast errors to quantify the uncertainty around their forecasts. For example, the Reserve Bank of Australia and the European Central Bank publish forecast fan charts with uncertainty bands derived from historical forecast errors. In the case of the Federal Reserve, to illustrate uncertainty surrounding the outlook, the Federal Open Market Committee's (FOMC) Summary of Economic Projections (SEP) provides a table of historical forecast root mean square errors (RMSEs). These are computed from the errors of several different forecasts, including, among others, the Survey of Professional Forecasters (SPF) and Congressional Budget Office (CBO). The historical RMSEs can provide a rough 70 percent confidence interval around a given forecast.

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. In practice, the historical accuracy measures published in the Federal Reserve's SEP are based on a 20-year rolling 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] 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] Justiniano and Primiceri (2008) and Diebold, Schorfheide,

[1]The forecasting literature builds on the initial work of Cogley and Sargent (2005) and Primiceri (2005) on VARs with stochastic volatility.

[2]Clark and Ravazzolo (2015) also consider VARs with GARCH and find that VARs with stochastic

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 or the CBO 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 FOMC's SEP, or the CBO, 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. 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 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) forecast 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

volatility yield more accurate forecasts.

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 with the full history of data for each variable (growth, unemployment, nominal short-term interest rate, inflation) 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 pseudo-real time estimates of forecast uncertainty and evaluate density forecasts implied by the SPF 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 forecasts (the approach of some central banks, noted above).

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 four or eight-quarter ahead forecasts at each survey date. 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).[3] Clements (2016) finds that density forecasts obtained from SPF histograms are no more accurate than density forecasts estimated from the his-

[3]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 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.

torical distributions of past point forecast errors.

The paper proceeds as follows. Section 2 describes the SPF forecasts and real time data used in evaluation. Section 3 presents our model of 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 or CPI, and the 3-month Treasury bill rate.[4] (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.[5]

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 1968:Q4 and ending in 2016:Q1 (however, forecasts for CPI inflation and the T-bill rate do not begin until 1981:Q3). 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.

Quantifying the forecast errors underlying our analysis requires a choice of outcomes against which to measure the forecasts. Sources such as Romer and Romer (2000), Sims (2002), and Croushore (2006) discuss various considerations for assessing the accuracy of

[4]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.

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

real-time forecasts. For our purposes, timeliness is a crucial consideration. We want to be able 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. For that purpose, as detailed below in the model exposition, we need a forecast error for the previous quarter $(t-1)$. For GDP growth and inflation in the GDP price index, we measure the quarter $t-1$ forecast error with the first (in time) estimate of the outcome, which is normally published by the Bureau of Economic Analysis at the end of the first month of quarter t. 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.[6] 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.

3 Model

We begin this section by setting the stage for our analysis with a conventional, simple 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.[7] For example, a simple time series model for a scalar variable y_t could take the form of an AR(1) specification with

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

[7] See, for example, Clark (2011) and D'Agostino, Gambetti, and Giannone (2013). These forecasting studies build on the seminal work of Cogley and Sargent (2005) and Primiceri (2005) on VARs with stochastic volatility.

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 $\log(\lambda_t)$.

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 have available 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 and the forecast is relative to observed information for period $t-1$. 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 $E_t y_t$, $E_t y_{t+1}$, $\ldots$, $E_t y_{t+H}$, as well as the forecasts made in previous periods. 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 an underlying 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 $t + h$:

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

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:

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

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 through $t-1$ and not t — adds a term to the decomposition that would not exist with textbook setups of time series models in which forecasts made at t reflect information through t.[8]

To obtain a useful econometric framework, we proceed to impose 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) 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.

Although we use Bayesian methods to measure forecast uncertainty as captured in the posterior distribution of forecast errors, at a conceptual level, 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

[8] 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): ${}_te_{t+h} = \sum_{i=1}^{h} \mu_{t+h|t+i}$.

forecast error as the sum of the (expected) variances of the individual terms in (1):

$$
\begin{aligned}
\text{var}_t(y_{t+h}) = \text{var}_t({}_t e_{t+h}) &= \text{var}_t({}_{t+h} e_{t+h}) + \sum_{i=1}^{h} \text{var}_t(\mu_{t+h|t+i}) \\
&= E_t\left[\text{var}_{t+h}({}_{t+h} e_{t+h})\right] + \sum_{i=2}^{h} E_t\left[\text{var}_{t+i-1}(\mu_{t+h|t+i})\right] + \text{var}_t(\mu_{t+h|t+1}). \quad (3)
\end{aligned}
$$

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

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

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 estimates of $\text{var}_t({}_t e_{t+h})$ from estimates of the conditional variances, for (1) the variance of the nowcast error, $\text{var}_t({}_t e_t)$, 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 multi-variate 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 $t-1$, but very little macroeconomic data on quarter t. Reflecting this timing, we 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.

More specifically, we define the data vector as:

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

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} and with the publication of actual data.

Our baseline model 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 both for correlation between level and scale shocks to $\boldsymbol{\eta}$:

$$\begin{aligned} \boldsymbol{\eta_t} &= \mathbf{A}\tilde{\boldsymbol{\eta}}_{\mathbf{t}} \qquad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ a_{H+1,1} & a_{H+1,2} & \dots & & 1 \end{bmatrix} \qquad (6) \\ \tilde{\boldsymbol{\eta}}_{\mathbf{t}} &= \boldsymbol{\Lambda_t}^{0.5}\boldsymbol{\epsilon_t}, \ \boldsymbol{\epsilon_t} \sim N(\mathbf{0}, \mathbf{I_{H+1}}), \ \boldsymbol{\Lambda_t} \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{H+1,t}) \\ \log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t}, \quad i = 1, \dots, H+1, \\ \boldsymbol{\nu_t} &\equiv (\nu_{1,t}, \nu_{2,t}, \dots, \nu_{H+1,t})' \sim \ N(\mathbf{0}, \boldsymbol{\Phi}). \end{aligned}$$

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. In practice, though, forecast errors from sources such as SPF sometimes go through substantial periods of bias, as documented in Croushore (2010). Accordingly, before estimating the model, we demean the elements of the data vector $\boldsymbol{\eta_t}$ using a pseudo-real time approach to computing a time-varying mean, with one-sided exponential smoothing.[9]

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 $\boldsymbol{\eta_t}$'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

[9]Specifically, let the raw data denote $\boldsymbol{\eta_t^O}$. We then compute the following exponentially weighted trend: $\boldsymbol{\bar{\eta}_t} = (1-\rho)\boldsymbol{\eta^O}_{\mathbf{t}} + \rho\boldsymbol{\bar{\eta}_{t-1}}$ where the scalar smoothing weight is set to 0.9 and the smoother is initialized with $\boldsymbol{\bar{\eta}_0} = \mathbf{0}$. In addition, when encountering missing data for the ith element of $\boldsymbol{\bar{\eta}_t}$, denoted $\bar{\eta}_t^i$, the time $t-1$ value of that ith element is rolled forward and we let $\bar{\eta}_t^i = \bar{\eta}_{t-1}^i$. Missing values inside an otherwise contiguous sequence of observations occur for a few forecast horizons of individual variables in the SPF. Finally, we construct $\boldsymbol{\eta_t} = \boldsymbol{\eta_t^O} - \boldsymbol{\bar{\eta}_t}$.

$\boldsymbol{\epsilon_t}$) that are Gaussian, but this does not imply the observed forecast errors and expectational updates to be Gaussian. The model allows the distributions of the observed expectational updates and forecast errors to feature fat tails. We discuss below a model extension to treat the conditional innovations $\boldsymbol{\epsilon_t}$ as having fat tails.

3.2.3 Estimating the model and forecast uncertainty

The model of (6) can be estimated by Bayesian Markov chain Monte Carlo methods (a Gibbs sampler) iterating over the following three blocks: First, taking estimates of $\boldsymbol{\Lambda_t}^{0.5}$ 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 $\tilde{\boldsymbol{\eta}}$.[10] 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

$$\log \tilde{\eta}_{i,t}^2 = \log \lambda_{i,t} + \log \epsilon_{i,t}^2, \;\; i = 1, \ldots, H+1, \tag{7}$$

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 $\tilde{\boldsymbol{\eta}}_{\mathbf{t}}$ are mutually uncorrelated (as opposed to $\boldsymbol{\eta}_{\mathbf{t}}$), the measurement equations (7) 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 responses in our sample. KSC and Primiceri

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

(2005) provide additional detail on the estimation algorithm.[11]

To estimate the uncertainty around multi-step forecasts, we simulate the posterior distribution of forecast errors using the model (6) and an approach like that detailed in Cogley, Morozov, and Sargent (2005). For each forecast horizon h, we need to simulate draws of the forecast error ${}_t e_{t+h}$, 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 70 percent coverage rates as the percentages of the observed forecast errors that fall within the 70 percent bands (credible sets) of the simulated forecast error distributions. In the next section, we detail these and the other forecast evaluation metrics considered.

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

3.2.4 An alternative model assuming constant variances

In light of common central bank practice (e.g., Reifschneider and Tulip 2007 and Table 2 of the Federal Reserve's SEP), the most natural benchmark against which to compare our proposed 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 considers forecast errors collected over the previous 20 calendar years). Accordingly, we report results obtained under a similar approach, where we will collect continuously updated estimates generated from rolling windows of data covering the most recent $R = 80$ quarterly observations. For simplicity, below we will refer to this specification as the "constant variance" approach, even though it acknowledges the potential for variance changes over time by using a rolling window of observations.

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:Q1–2016:Q1 for GDP growth, unemployment, and GDP inflation; and 1996:Q4–2016:Q1 for CPI inflation and the T-bill rate. As we detail in the robustness results below, our main findings apply to rolling windows shorter or longer than the baseline.

The constant variance model simply assumes 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):

$$\boldsymbol{\eta_t} \sim N(\mathbf{0}, \boldsymbol{\Sigma}). \tag{8}$$

For the sake of comparing not only point forecasts but also predictive densities, we also employ Bayesian methods to estimate this model within the quasi-real-time setup describe above. For this analysis we assume a diffuse inverse-Wishart prior.[12]

[12] The occurrence of missing observations for some SPF responses is handled with a simple Gibbs sampler,

4 Evaluation metrics

The previous section described two alternative volatility models — our proposed stochastic volatility model and an alternative constant-variance benchmark. This section considers two measures of density forecast accuracy to assess the 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 70 percent prediction interval, as well as some selected empirical quantiles of the forecast distribution.[13] A frequency of more (less) than 70 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.[14]

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.[15] 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

$$CRPS_t(y^o_{t+h}) = \int_{-\infty}^{\infty} \left(F(z) - 1\{y^o_{t+h} \leq z\}\right)^2 dz = E_f|Y_{t+h} - y^o_{t+h}| - 0.5 E_f|Y_{t+h} - Y'_{t+h}|, \quad (9)$$

where F denotes the cumulative distribution function associated with the predictive density

which iterates between draws from the inverse Wishart posterior for $\mathbf{\Sigma}|(\boldsymbol{\eta}_\mathbf{t} \forall\, t = 1, \ldots, T)$ and draws of $\boldsymbol{\eta}_\mathbf{t}|\mathbf{\Sigma}$ that are used for padding the missing observations. The Gibbs sampler is initialized with a value for $\mathbf{\Sigma}$ drawn from an inverse Wishart with mean equal to the identity matrix and $H + 3$ degrees of freedom (the minimum degrees of freedom necessary to ensure the existence of the first moment for an inverse Wishart of size $H + 1$).

[13] We also considered analyzing results for the 90 percent prediction interval. However, with more observations available at the 70 percent level than the 90 percent level, we chose to devote attention to the former rather than the latter.

[14] 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.

[15] recent applications can be found, for example, in Ravazolo and Vahey (2014) as well as Clark and Ravazzolo (2015).

f, $1\{y^o_{t+h} \leq z\}$ denotes an indicator function taking value 1 if the outcome $y^o_{t+h} \leq z$ and 0 otherwise, and Y_{t+h} and Y'_{t+h} are independent random draws from the posterior predictive density. To provide a rough gauge of whether the differences in density forecast accuracy — for the baseline constant variance approach and our proposed stochastic volatility approach — are significant, we apply Diebold and Mariano (1995) t-tests for equality of the average loss as measured by the CRPS.[16] We compute the CRPS by integrating over the empirical CDF generated from draws simulated from the predictive density of our multivariate stochastic volatility model.[17]

As noted above, a number of studies have compared the density forecast performance 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, 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. The section concludes with a discussion of directions in which the model could be extended.

[16] Amisano and Giacomini (2007) extend the results of Giacomini and White (2006) for point forecasts to density forecasts, developing a set of weighted likelihood ratio tests. In our application to tests for density forecasts, we do not employ a weighting scheme (or put another way, we use equal weights).

[17] Specifically, we compute $\widehat{CRPS}_t(y^o_{t+h}) = \sum_{i=1}^{N} \left(\hat{F}(z_i) - 1\{y^o_{t+h} \leq z_i\}\right)^2$, by summing over draws $i = 1, 2, \ldots, N$ and denoting the empirical CDF of a draw with value z_i by $\hat{F}(z_t)$.

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. The results we highlight for GDP growth and unemployment also apply to the other variables. The charts report both the raw data of $\boldsymbol{\eta_t}$ (red lines) as well as the one-sided estimates of the means (black dotted lines) removed before model estimation.

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. To get a sense of whether the serial correlation in the expectational updates is significant enough to warrant modeling, for each variable we estimated vector autoregressions based on the vector $\boldsymbol{\eta_t}$ using 0 to 4 lags. In all but one case, BIC indicates the optimal lag order to be 0 (the exception is the T-bill rate, for which the BIC selected one lag). By this measure, the serial correlation in the expectational updates appears modest enough to be consistent with our proposed multivariate stochastic volatility specification.

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. Finally, the one-sided estimates of means suggest some low-frequency biases (again, consistent with the results of Croushore 2010), which the mean estimates seek to remove before model estimation. For GDP growth, the mean estimates vary in sign at all horizons, although they are sizably negative around the time of the recessions in the early 1980s. The time-varying mean estimates for the unemployment rate have a noticeable cyclicality, rising around recessions. As indicated in the appendix, the mean estimates for inflation are fairly consistently negative from about 1980 until the early 2000s, and the mean estimates for the T-bill rate tend to fall around recessions and recover over economic 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 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; 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 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. However, the cyclical pattern appears smaller for inflation, except that CPI inflation spiked sharply with the Great Recession, presumably due to the dramatic, unexpected falloff in inflation that occurred.

- 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 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 Crisis-period 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 Crisis-period 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 (demeaned in real time using the one-sided ES filter) for each variable along with 70 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. (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. Tables 1 through 3 quantify empirical coverage rates. In the discussion below, we focus on 70 percent coverage rates, because there are many fewer observations available for evaluating accuracy further out in the tails of the distributions.

The charts of the time paths of 70 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 sample dates back to 1985), 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 approach varies more than does the width of intervals based on constant variances. For GDP growth, unemployment, and GDP inflation, the SV-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-based estimates, the similarities across horizons are particularly strong for horizons 1 through 4 (omitting the nowcast horizon).[18] 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.[19] More broadly, with these variance estimates reflecting

[18] 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.

[19] In unreported results, we have used the estimates of Φ at each forecast origin to construct the correlation

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 quantify the accuracy of the 70 percent confidence intervals shown in Figures 8-12. These show the intervals based on our stochastic volatility model to be consistently more accurate than the intervals based on the constant variance approach. Although we cannot claim that the SV-based approach yields correct coverage in all cases, it typically improves on the alternative approach, which often yields coverage rates above 70 percent, reflecting bands that are too wide. For example, for GDP growth, the SV-based coverage rates range (across horizons) from 68 percent to 74 percent, with no departures from 70 percent large enough to be statistically significant, whereas the constant variance-based rates range from 77 percent to 81 percent, with three departures from 70 percent large enough to be statistically significant. For the T-bill rate, the SV-based rates are lower than the constant-variance-based rates at forecast horizons of 2 quarters or less — e.g., at the 2-step horizon, 72 percent with SV versus 80 percent for the constant variance. Admittedly, however, accurate coverage is somewhat more difficult to achieve for inflation. For GDP inflation (over the 1984-2016 period), both approaches yield coverage rates that exceed 70 percent, with the SV rates better at shorter horizons but not longer horizons. For CPI inflation (over the 1996 to 2016 period), both approaches yield coverage rates that, at most horizons, are modestly below 70 percent.

To quantify the accuracy of the lower and upper bounds of the confidence intervals, Tables 2 and 3 report one-sided coverage rates associated with nominal quantiles of 15.87 and 84.13 percent, respectively. In both cases, the SV-based estimates are generally more accurate than the estimates based on constant variances. In the case of the lower quantile, under the constant variance approach, the empirical quantiles are almost always below the nominal rates of 15.87 percent, with several observations significantly different from the nominal quantile. The quantiles obtained with our SV-based model are consistently higher, and therefore usually more accurate, than those obtained with the constant variance approach. In the case of the upper quantile, the SV-based estimates are closer to the nominal rate of 84.13 percent than are the constant-variance based estimates in most variable-horizon combinations.

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.

To provide a broader assessment of density forecast accuracy, Table 4 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. For all variables except CPI inflation, our SV model consistently offers density accuracy gains over the constant variance specification. The gains are on the order of 10 percent for GDP growth, GDP inflation, and the T-bill rate and smaller for the unemployment rate.[20] In the unique case of CPI inflation, the density forecasts obtained with our SV model are worse than those obtained with a constant variance approach. This pattern may be driven by the shorter sample available for CPI forecasts (particularly for model estimation), as well as the influence of the large swings in inflation observed around the Great Recession. Some of the model extensions we describe later in this section could mitigate this challenge and improve the performance of the SV model for CPI inflation.

5.3 Out-of-Sample Forecasts: Robustness

To ensure that our broad forecast findings presented above are robust, we have conducted a variety of checks of our specification choices. These checks include omitting the one-sided demeaning of expectational updates and forecast errors that we deployed in the baseline case and making the rolling window underlying the constant variance specification either shorter or longer.[21] In this subsection, we briefly summarize these robustness checks, with some details provided in the supplementary appendix.

We first verify that our one-sided demeaning of the data before model estimation and forecast evaluation does not drive the baseline results. Tables 5 and 6 provide 70 percent coverage rates and CRPS results obtained when we do not demean the data, and just use the raw expectational updates and forecast errors for model estimation and forecast evaluation. Our main baseline findings are evident in these estimates. The one-sided demeaning seems to help coverage and CRPS accuracy in some cases and harm it in others, without any consistent pattern or really sizable consequences. Without any attempt to demean the

[20] 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.

[21] In unreported results, we also considered preserving our one-sided demeaning before model estimation and ex post, full-sample demeaning of the forecast errors before evaluation. These results are very similar to those for the baseline, because the one-sided demeaning makes the forecast errors close enough to 0 that the ex post demeaning after evaluation has little additional effect.

data, it remains the case that the intervals based on our stochastic volatility model are consistently more accurate than the intervals based on the constant variance approach, and that, for all variables except CPI inflation, our SV model consistently offers modest to sizable gains — over the constant variance specification — in density accuracy as measured by the CRPS.

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).[22] 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. Nonetheless, as detailed in appendix Tables 1 and 2, 70 percent coverage and CRPS results for the shorter sample with the longer rolling window used to compute the constant variance are very similar to those of the baseline. To the extent notable differences occur in the coverage rates (versus the baseline), they occur more for CPI inflation and the T-bill rate (for which the evaluation sample covers only about 15 years with the longer rolling window) than the other variables. For density accuracy (SV versus constant variance) as measured by the CRPS, results for the shorter sample with the longer rolling window are broadly similar to those of the baseline. In fact, for the most part, the gains to be achieved with SV are modestly larger when the constant variance is based on a rolling window of 80 observations than the baseline window of 60 observations. That said, the results on density accuracy for the CPI are worse with the longer rolling window than with the baseline.

Other alternatives shorten the rolling window to 40 or 20 observations. With these window settings, we could lengthen the evaluation sample, but to facilitate comparisons to the baseline case, we leave the evaluation sample unchanged from the baseline. As detailed in appendix Tables 3 and 4, results on coverage and CRPS accuracy with a 40 year rolling window used for the constant variance case are very similar to those reported for the baseline. Broadly, shortening the rolling window from 60 to 40 observations doesn't have much effect. With the somewhat shorter rolling window, it remains the case that the

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

intervals based on our stochastic volatility model are consistently more accurate than the intervals based on the constant variance approach, and that, for all variables except CPI inflation, our SV model consistently offers modest to sizable gains — over the constant variance specification — in density accuracy. The same basic finding applies when the rolling window is shortened to just 20 observations, as detailed in appendix Tables 5 and 6. Shortening the constant variance window helps in some cases and hurts in others, but our main findings of a payoff — in 70 percent coverage accuracy and CRPS-measured accuracy — still apply.

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.

Second, the model could be extended to make use of forecasts from multiple sources. In Reifschneider and Tulip (2007) 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.[23]

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

[23]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.

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 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

[To be written]

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.

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

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

Carriero, A., T. E. Clark, and M. Marcellino (2016), "Common drifting volatility in large Bayesian VARs," *Journal of Business and Economic Statistics* 34, 375-390.

Chiu, C.W., Mumtaz, H., and Pinter, G. (2015), "Forecasting with VAR models: Fat Tails and Stochastic Volatility," Bank of England Working Paper No. 528.

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

Cogley, T., Morozov, S., and Sargent, T. J. (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.

Croushore D. 2006. Forecasting with real-time macroeconomic data. In *Handbook of Economic Forecasting*, Elliott G, Granger C, Timmermann A (eds). North Holland: Amsterdam.

Croushore D. 2010. "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data," *BE Journal of Macroeconomics: Contributions* 10, Article 10.

Croushore D, Stark T. 2001. A real-time data set for macroeconomists. *Journal of Econometrics* **105**: 111-130.

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

D'Agostino A, Gambetti L, Giannone D. 2013. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics* **28**: 82-101.

Del Negro M, Primiceri G. (2015), Time varying structural vector autoregressions and monetary policy: A corrigendum. *Review of Economic Studies*, 82, 1342-1345.

Diebold F, Mariano, R. 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* **13**: 253-263.

Diebold, Francis X., Frank Schorfheide, and Minchul Shin (2016), "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," manuscript, University of Pennsylvania.

Durbin, J. and S. J. Koopman (2002, September). "A simple and efficient simulation smoother for state space time series analysis." *Biometrika 89*(3), 603–615.

Giacomini R, White H. 2006. Tests of conditional predictive ability. *Econometrica* **74**: 1545-1578.

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

Gneiting T, Raftery A. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* **102**: 359-378.

Gneiting T, Ranjan R. 2011. Comparing density forecasts using threshold and quantile weighted proper scoring rules. *Journal of Business and Economic Statistics* **29**: 411-422.

Jacquier E., Polson N., and Rossi P. (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.

Kim S, Shephard N, Chib S. 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* **65**: 361-393.

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

Kozicki, S. and B. Hoffman (2004). Rounding Error: A Distorting Influence on Index Data. *Journal of Money, Credit, and Banking 36*. 319-38.

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.

Primiceri, G. 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.

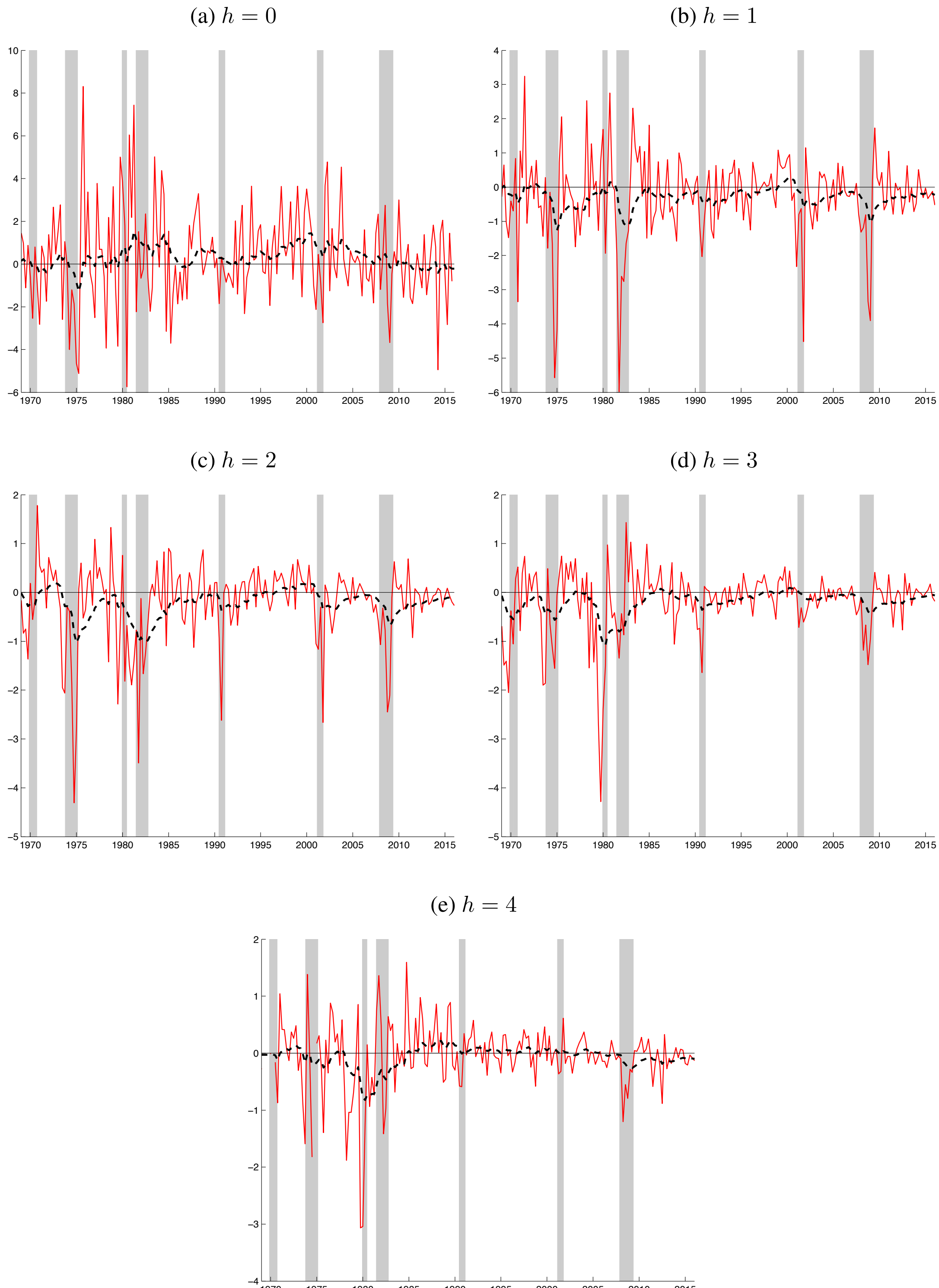
Romer C, Romer D. 2000. Federal Reserve information and the behavior of interest rates. *American Economic Review* **90**: 429-457.

Sims C. 2002. The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity* **2**: 1-40.

Stock J., and Watson, M. (2016), "Core Inflation and Trend Inflation," *Review of Economics and Statistics* 98, xxx.

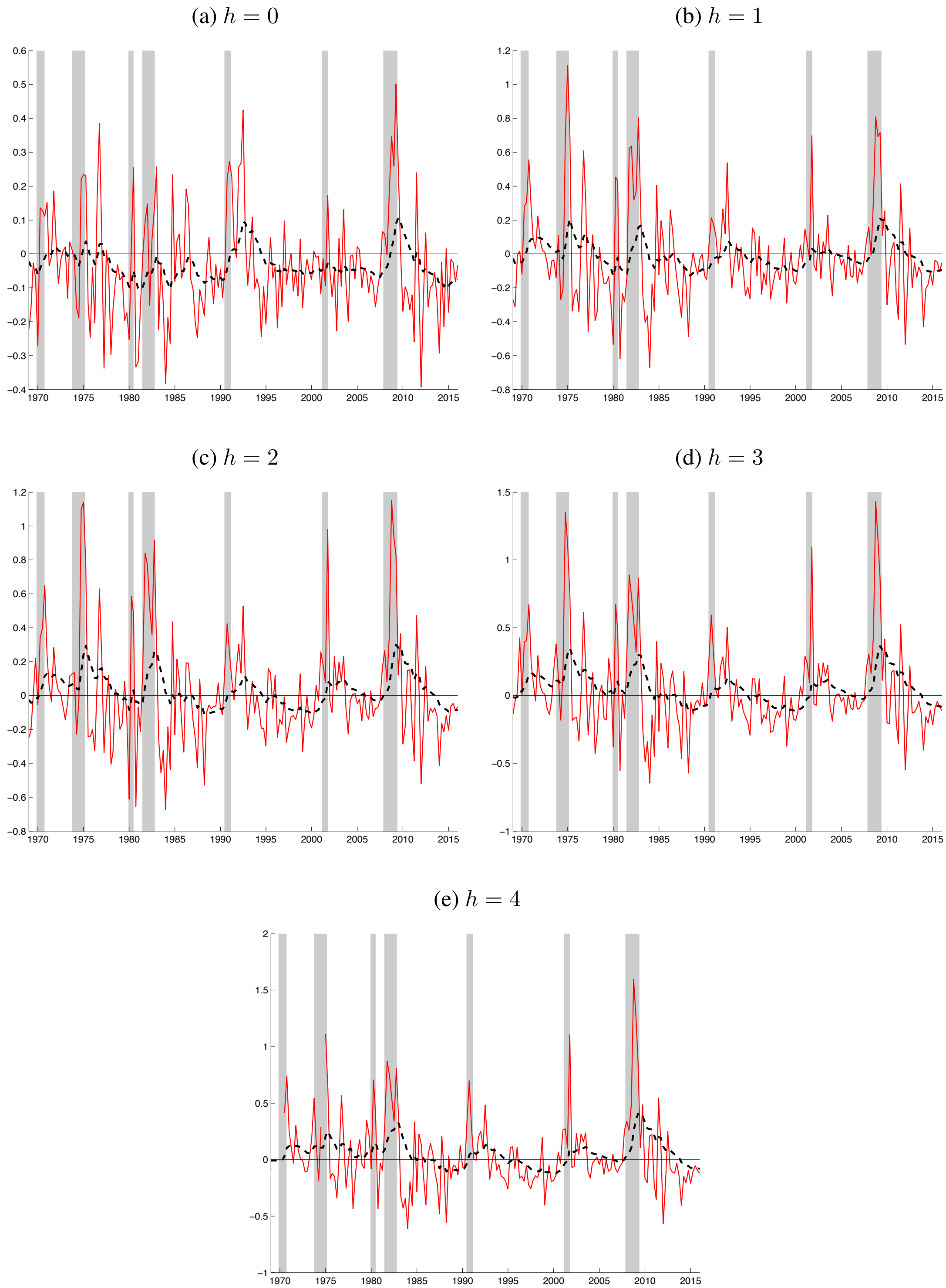
Stock, J.H. & Watson, M.W. (2007). Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking* 39, 3-33.

Figure 1: Expectational Updates for Real GDP Growth



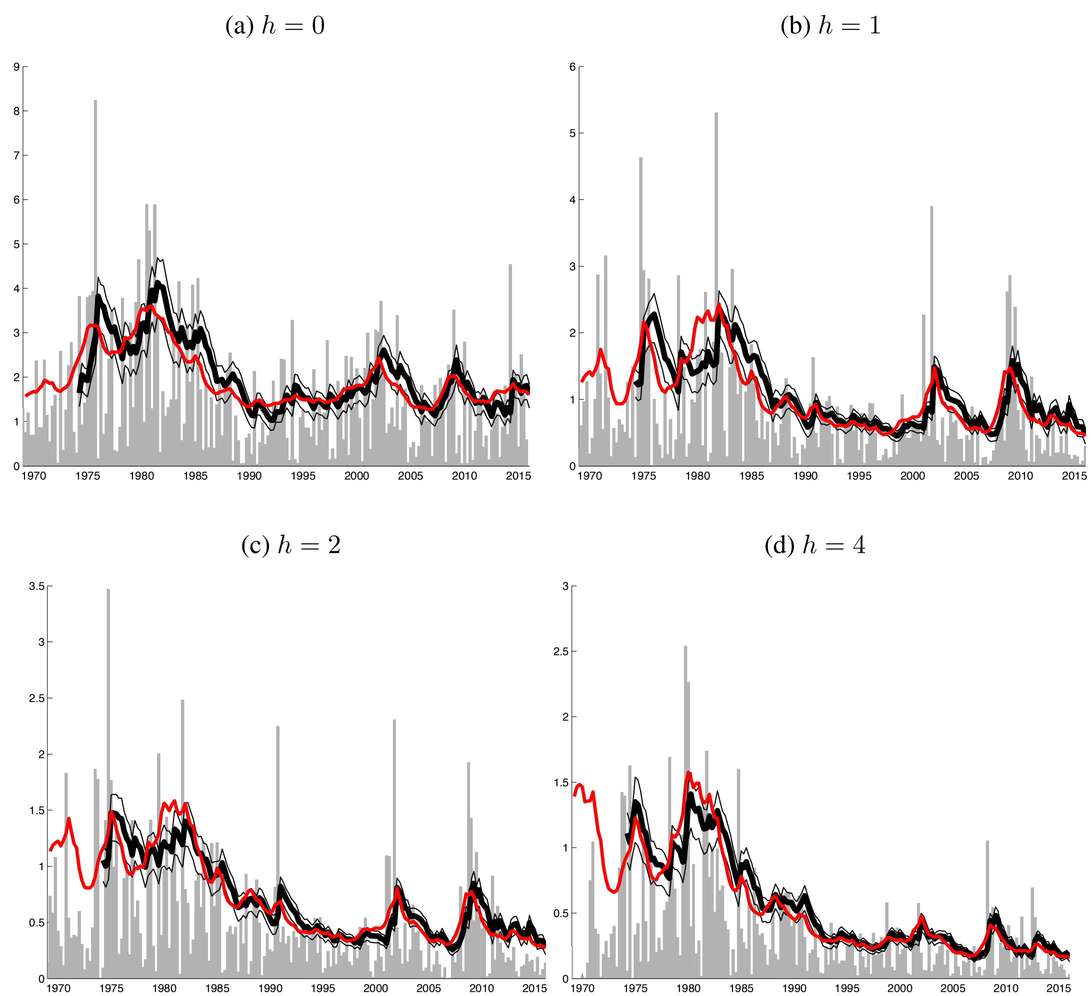
Note: The figure reports (in the red lines) the elements of the vector of expectational updates η_t used in model estimation, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 2: Expectational Updates for the Unemployment Rate



Note: The figure reports (in the red lines) the elements of the vector of expectational updates η_t used in model estimation, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

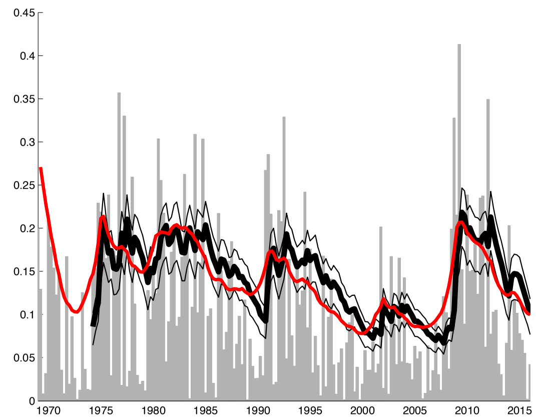
Figure 3: Stochastic Volatility in Expectational Updates for Real GDP Growth



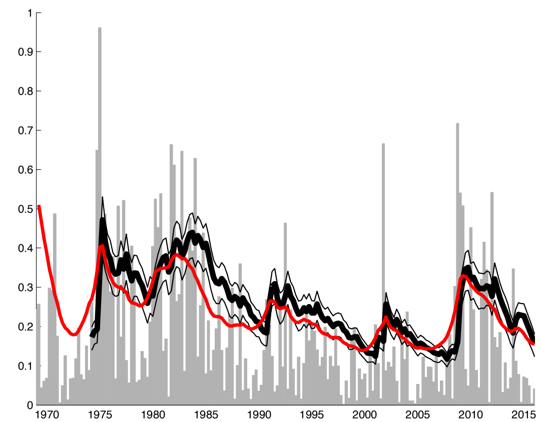
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond 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 4: Stochastic Volatility in Expectational Updates for the Unemployment Rate

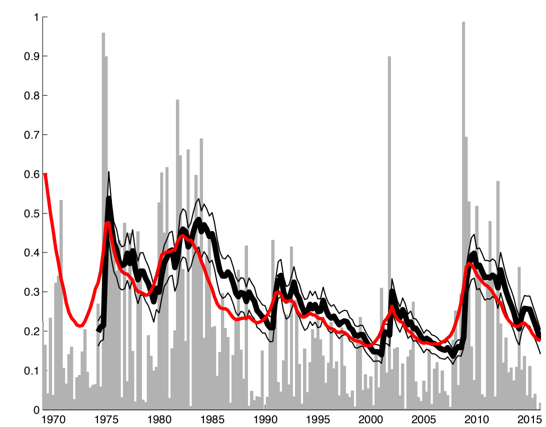
(a) $h = 0$



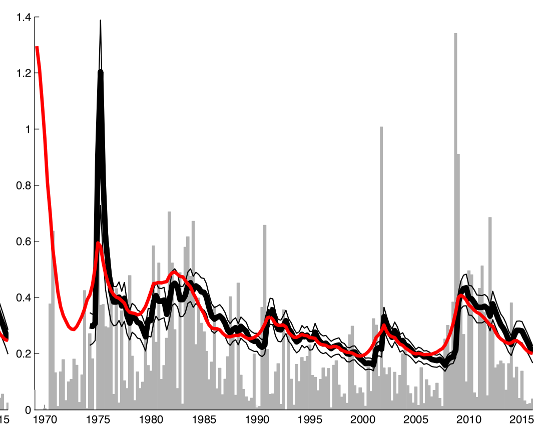
(b) $h = 1$



(c) $h = 2$



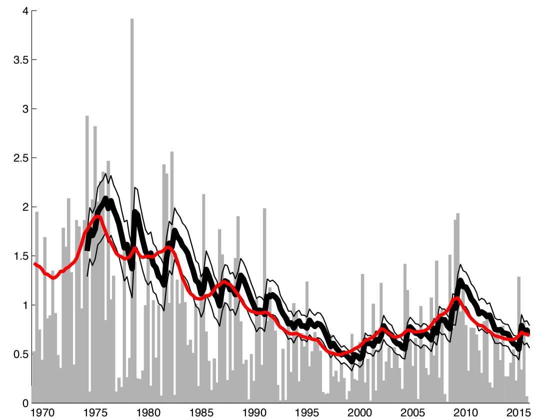
(d) $h = 4$



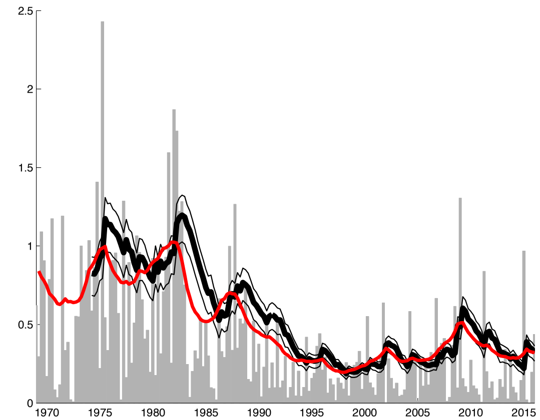
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond 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.

Figure 5: Stochastic Volatility in Expectational Updates for GDP Deflator Inflation

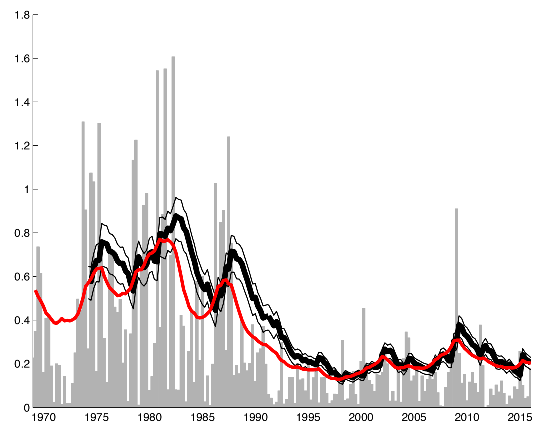
(a) $h = 0$



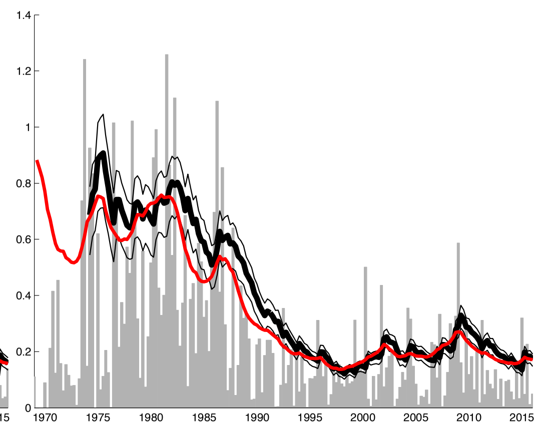
(b) $h = 1$



(c) $h = 2$



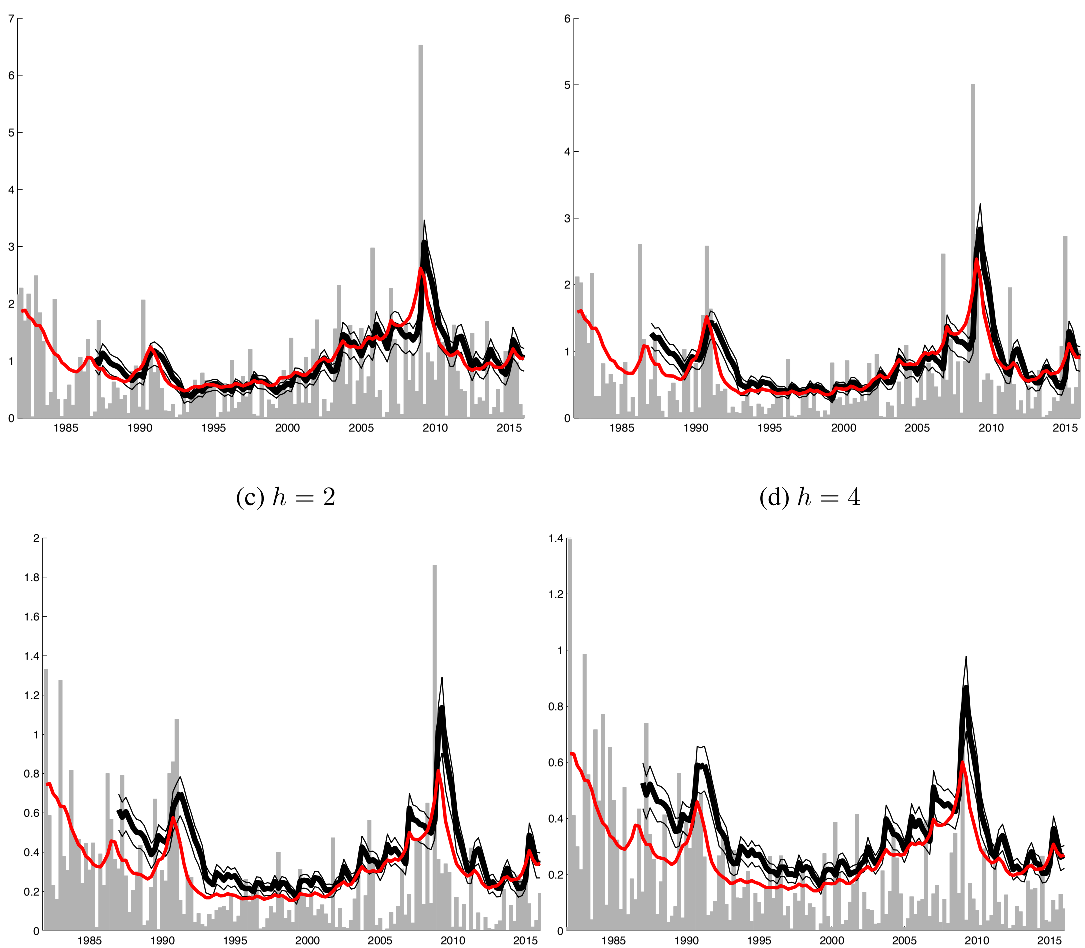
(d) $h = 4$



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.

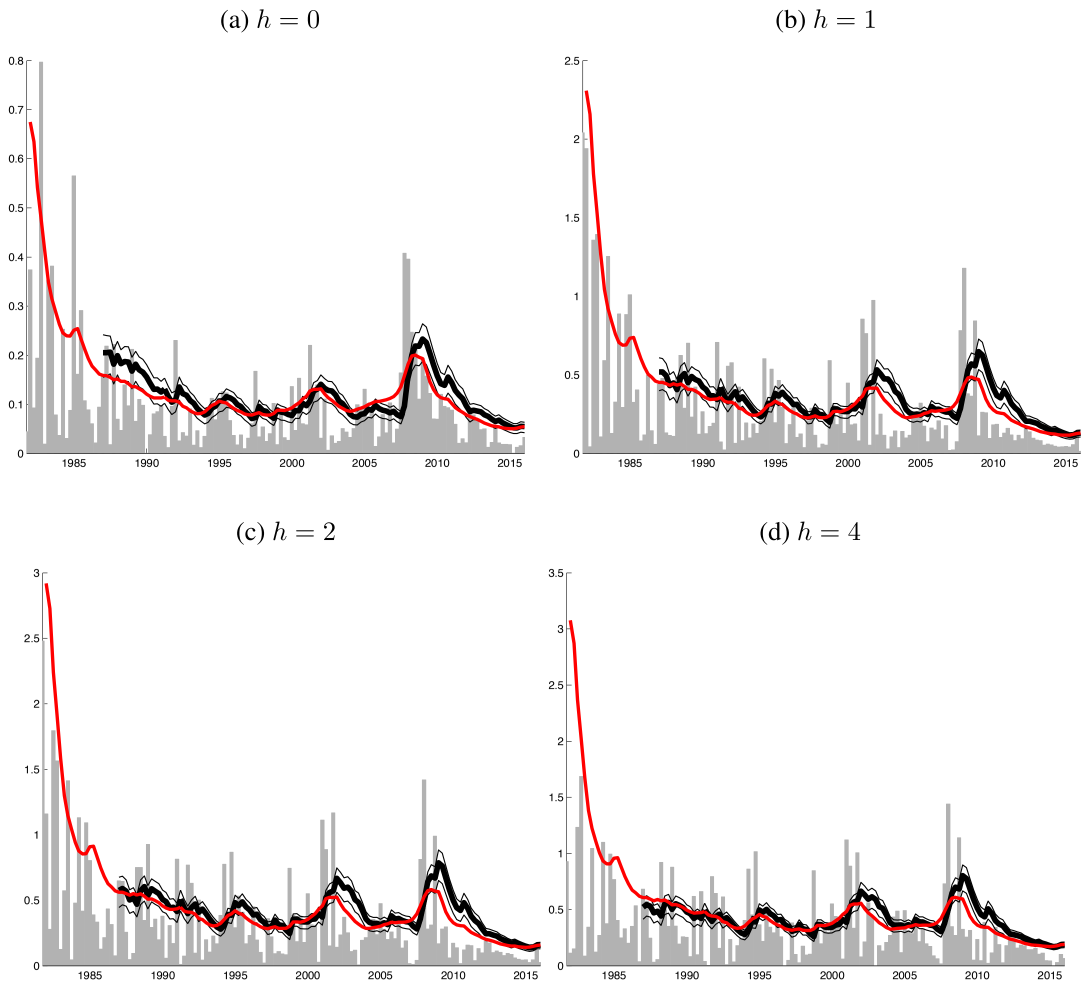
Figure 6: Stochastic Volatility in Expectational Updates for CPI Inflation

(a) $h = 0$ (b) $h = 1$ (c) $h = 2$ (d) $h = 4$



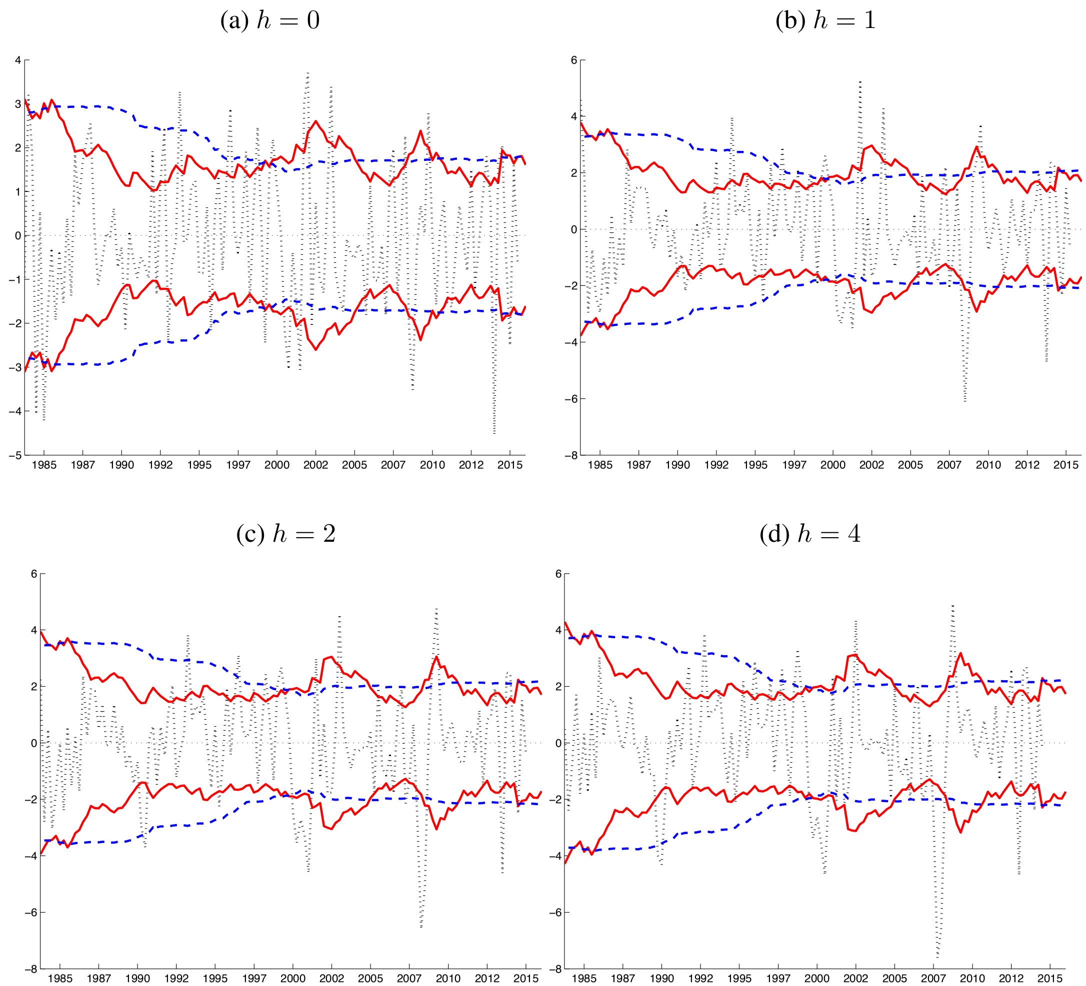
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond 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.

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



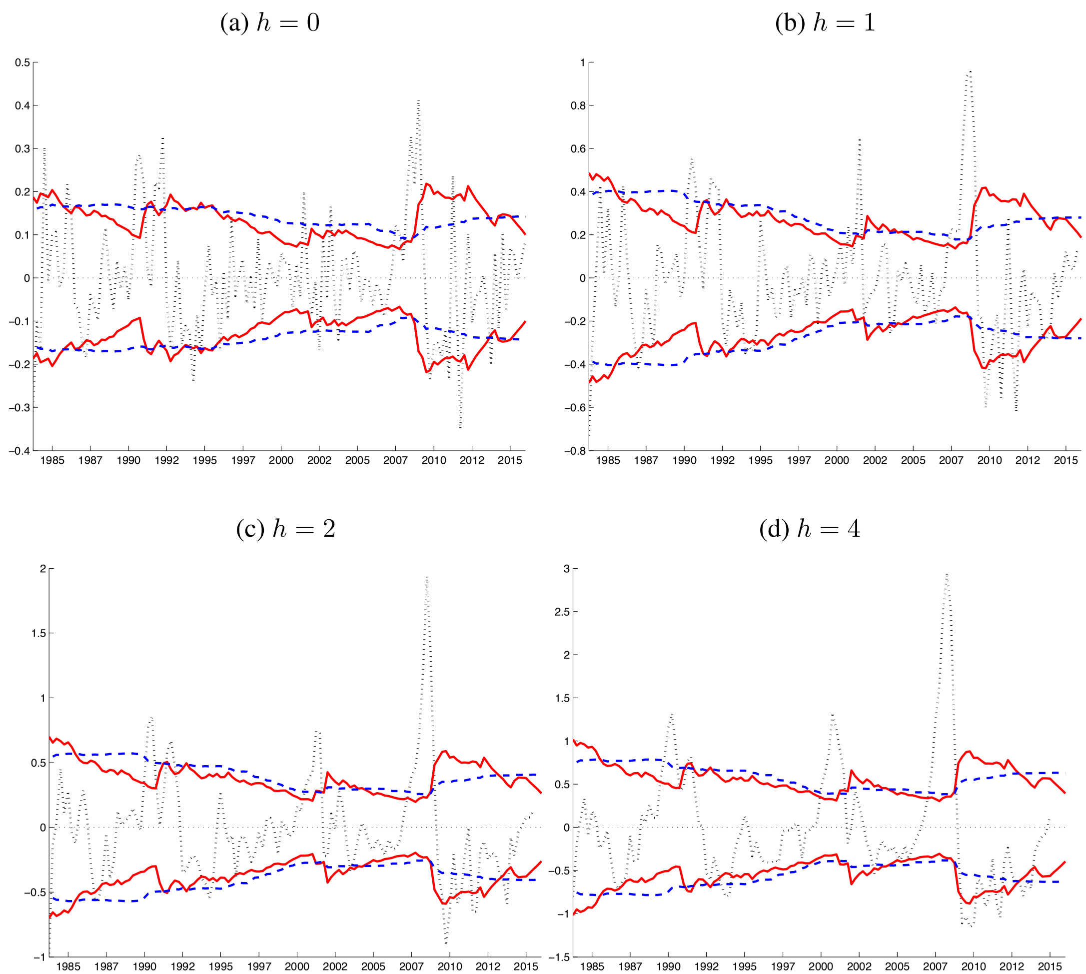
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond 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



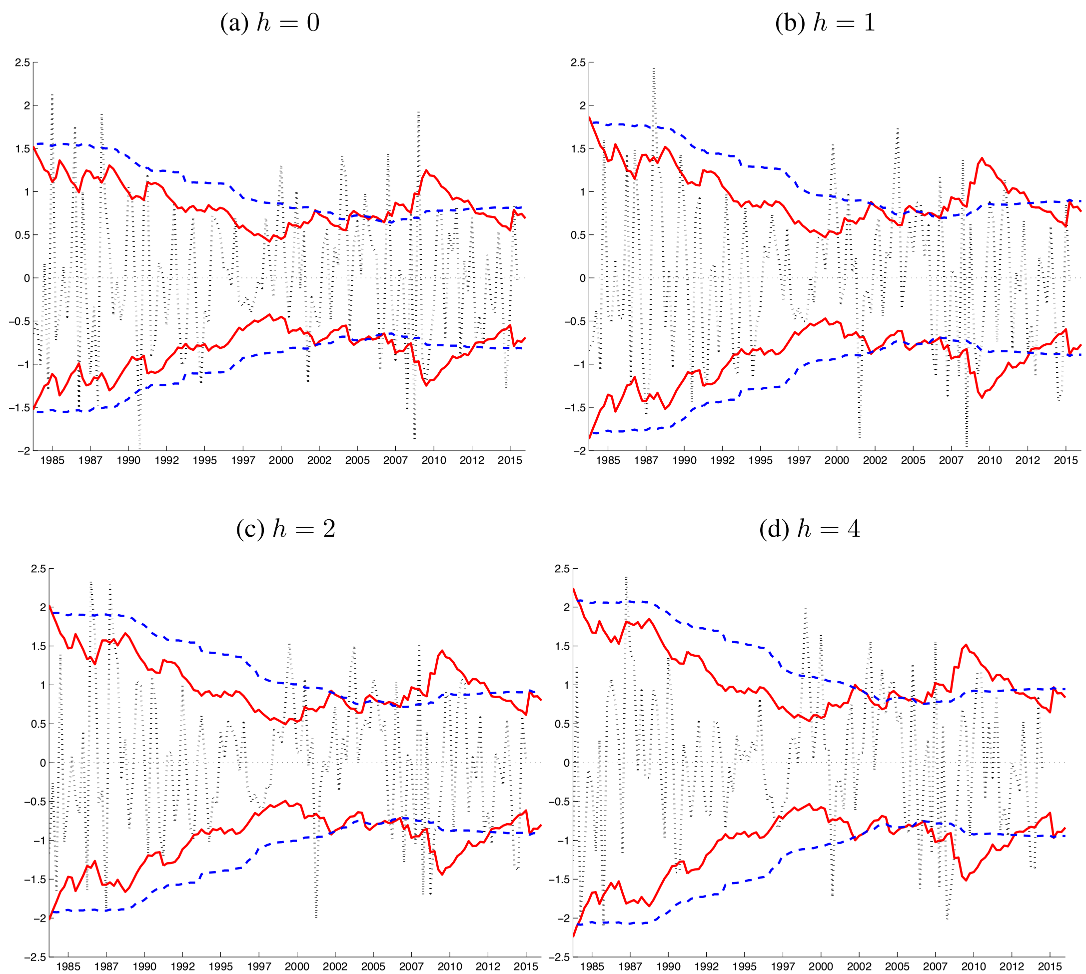
Note: Each figure reports forecast errors (demeaned in real time using the one-sided ES filter) for each variable along with 70 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained from our stochastic volatility model.

Figure 9: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for the Unemployment Rate



Note: Each figure reports forecast errors (demeaned in real time using the one-sided ES filter) for each variable along with 70 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.

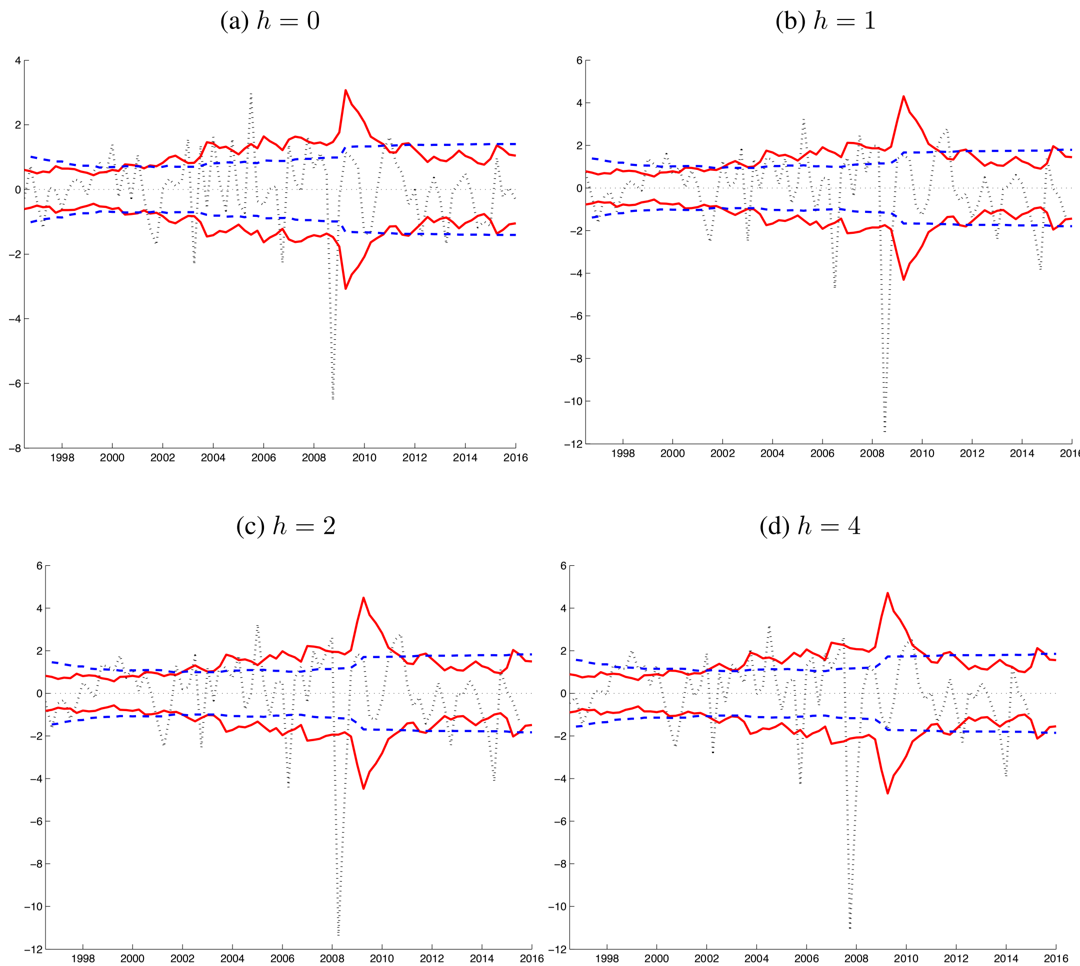
Figure 10: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for GDP Deflator Inflation



(a) $h = 0$ (b) $h = 1$ (c) $h = 2$ (d) $h = 4$

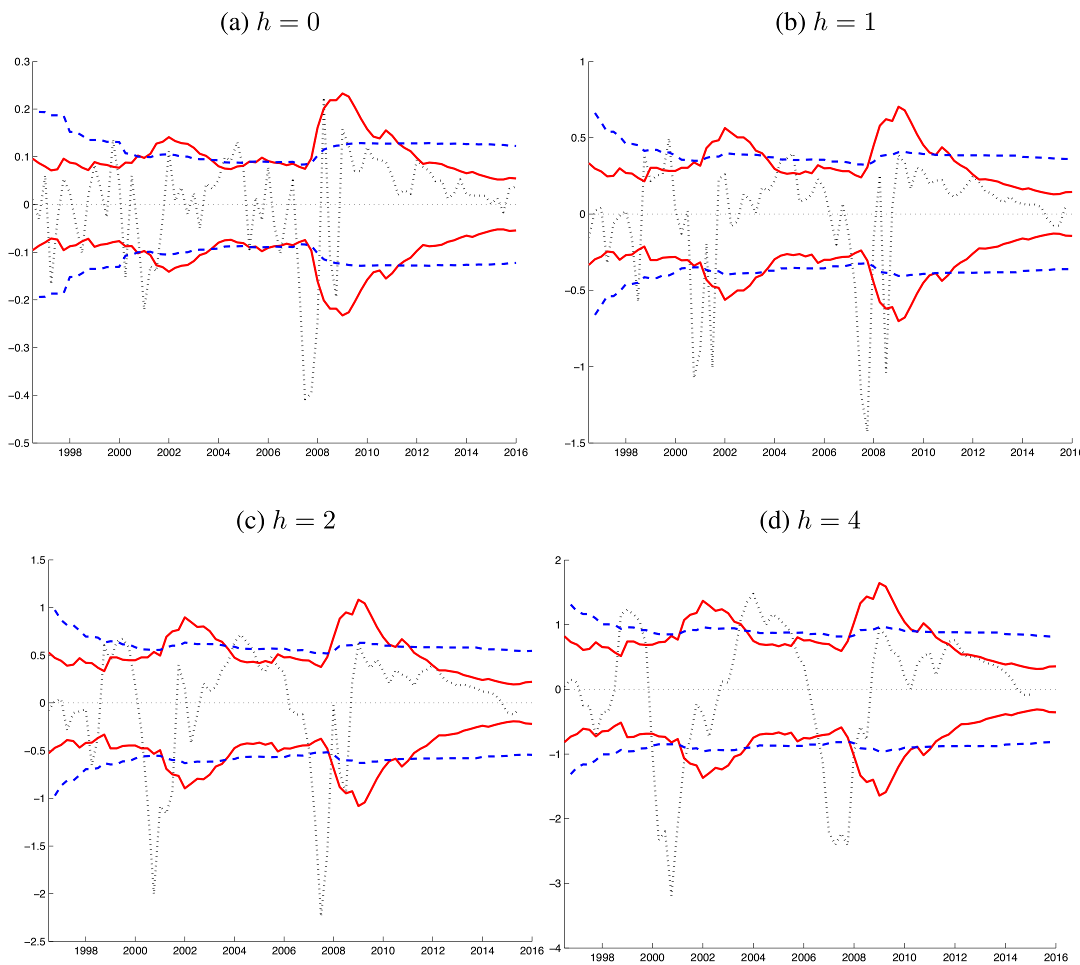
Note: Each figure reports forecast errors (demeaned in real time using the one-sided ES filter) for each variable along with 70 percent confidence intervals, one set (in blue) obtained with the constant variance approach and a 60 observation rolling window of observations and the other (in red) obtained 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 (demeaned in real time using the one-sided ES filter) for each variable along with 70 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



Note: Each figure reports forecast errors (demeaned in real time using the one-sided ES filter) for each variable along with 70 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.

Table 1: 70 percent forecast error coverage rates

Variable	SPF horizon 0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	67.72	71.43	74.40	72.58	69.11	1984:Q1
UNRATE	75.19	73.44	72.44	73.02	68.00	1984:Q1
PGDP	74.02	73.02	78.40**	78.23**	78.05**	1984:Q1
CPI	70.51	66.23	64.47	65.33	62.16	1996:Q4
TBILL	78.21*	81.82**	72.37	65.33	63.51	1996:Q4
Panel B: CONST						
RGDP	77.17**	80.95***	80.80**	76.61	77.24	1984:Q1
UNRATE	74.42	75.78	77.17	73.02	69.60	1984:Q1
PGDP	80.31***	82.54***	79.20**	80.65***	78.86**	1984:Q1
CPI	71.79	63.64	63.16	64.00	64.86	1996:Q4
TBILL	76.92	85.71***	80.26*	72.00	66.22	1996:Q4

Note: The table reports the empirical coverage rates of 70 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. Statistically significant departures from correct nominal coverage are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 2: Coverage rate of lower quantile

15.87% quantile						
	SPF horizon					
Variable	0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	15.75	16.67	13.60	14.52	16.26	1984:Q1
UNRATE	10.08**	12.50	12.60	14.29	15.20	1984:Q1
PGDP	12.60	12.70	12.00	9.68*	11.38	1984:Q1
CPI	14.10	14.29	15.79	14.67	16.22	1996:Q4
TBILL	15.38	11.69	15.79	14.67	17.57	1996:Q4
Panel B: CONST						
RGDP	12.60	9.52*	9.60	12.10	13.82	1984:Q1
UNRATE	11.63	9.38**	8.66*	12.70	14.40	1984:Q1
PGDP	9.45**	9.52**	10.40	8.06**	10.57	1984:Q1
CPI	12.82	14.29	11.84	12.00	12.16	1996:Q4
TBILL	14.10	10.39	13.16	16.00	16.22	1996:Q4

Note: The table reports the coverage rate of the 15.87% quantile of the out-of-sample forecast distribution. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. Statistically significant departures from correct nominal coverage are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 3: Coverage rate of upper quantile

84.13% quantile						
	SPF horizon					
Variable	0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	80.31	86.51	85.60	83.87	84.55	1984:Q1
UNRATE	84.50	85.94	85.04	85.71	82.40	1984:Q1
PGDP	85.04	84.92	89.60*	87.10	89.43	1984:Q1
CPI	83.33	80.52	80.26	78.67	78.38	1996:Q4
TBILL	92.31**	90.91*	85.53	78.67	75.68	1996:Q4
Panel B: CONST						
RGDP	88.98*	90.48**	90.40**	88.71	90.24**	1984:Q1
UNRATE	86.05	85.16	86.61	85.71	84.00	1984:Q1
PGDP	89.76**	91.27***	89.60**	88.71	89.43	1984:Q1
CPI	83.33	77.92	75.00	76.00	77.03	1996:Q4
TBILL	91.03*	96.10***	93.42**	88.00	82.43	1996:Q4

Note: The table reports the coverage rate of the 84.13% quantile of the out-of-sample forecast distribution. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. 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

Variable		SPF horizon 0	1	2	3	4	eval. begin
RGDP	(SV rel.)	9.96%	10.83%	10.65%	10.07%	9.73%	1984:Q1
	(CONST abs.)	2.13	2.41	2.51	2.59	2.65	
UNRATE	(SV rel.)	3.03%	3.19%	2.51%	1.62%	−0.03%	1984:Q1
	(CONST abs.)	0.15	0.31	0.44	0.57	0.68	
PGDP	(SV rel.)	9.43%	9.90%	11.06%	10.66%	10.57%	1984:Q1
	(CONST abs.)	1.03	1.14	1.20	1.23	1.29	
CPI	(SV rel.)	−5.54%	−8.61%	−5.24%	−5.86%	−7.06%	1996:Q4
	(CONST abs.)	1.20	1.62	1.75	1.76	1.77	
TBILL	(SV rel.)	9.16%	6.74%	9.46%	7.42%	5.49%	1996:Q4
	(CONST abs.)	0.13	0.40	0.66	0.89	1.09	

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 2016:Q1 (evaluated against realized data as far as available). 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. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing.

Table 5: Coverage rates, forecast errors and updates not demeaned

	SPF horizon					
Variable	0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	66.93	70.63	72.00	71.77	70.73	1984:Q1
UNRATE	72.87	73.44	68.50	65.08	64.80	1984:Q1
PGDP	72.44	71.43	72.80	68.55	68.29	1984:Q1
CPI	66.67	63.64	57.89*	66.67	63.51	1996:Q4
TBILL	80.77**	79.22*	73.68	65.33	54.05	1996:Q4
Panel B: CONST						
RGDP	80.31***	83.33***	82.40***	81.45***	81.30***	1984:Q1
UNRATE	74.42	78.12**	74.80	74.60	68.00	1984:Q1
PGDP	77.95**	80.95***	79.20***	78.23**	74.80	1984:Q1
CPI	71.79	58.44	56.58*	57.33*	60.81	1996:Q4
TBILL	79.49**	84.42***	80.26	77.33	67.57	1996:Q4

Note: The table reports the empirical coverage rates of 70 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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 6: CRPS accuracy, forecast errors and updates not demeaned

		SPF horizon					
Variable		0	1	2	3	4	eval. begin
RGDP	(SV rel.)	10.63%	11.90%	11.97%	11.29%	10.57%	1984:Q1
	(CONST abs.)	2.35	2.64	2.78	2.88	2.95	
UNRATE	(SV rel.)	1.63%	2.03%	1.54%	−0.21%	−2.13%	1984:Q1
	(CONST abs.)	0.17	0.34	0.49	0.63	0.76	
PGDP	(SV rel.)	9.63%	9.63%	10.13%	9.41%	9.06%	1984:Q1
	(CONST abs.)	1.13	1.27	1.35	1.41	1.49	
CPI	(SV rel.)	−5.92%	−4.66%	−5.61%	−10.59%	−8.03%	1996:Q4
	(CONST abs.)	1.30	1.87	1.95	1.88	1.98	
TBILL	(SV rel.)	8.30%	13.84%	8.27%	10.49%	8.61%	1996:Q4
	(CONST abs.)	0.16	0.49	0.76	1.10	1.38	

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 2016:Q1 (evaluated against realized data as far as available). 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 1: 70 percent forecast error coverage rates, 80 quarter rolling window for constant variance, shorter sample

	SPF horizon					
Variable	0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	66.36	66.98	69.52	67.31	63.11	1989:Q1
UNRATE	75.23	72.22	69.16	69.81	62.86	1989:Q1
PGDP	74.77	73.58	79.05**	77.88*	75.73	1989:Q1
CPI	72.41	70.18	69.64	72.73	70.37	2001:Q4
TBILL	82.76**	87.72***	80.36	72.73	68.52	2001:Q4
Panel B: CONST						
RGDP	77.57**	75.47	78.10*	75.96	78.64*	1989:Q1
UNRATE	74.31	76.85	75.70	73.58	64.76	1989:Q1
PGDP	79.44**	81.13***	80.00**	82.69***	79.61**	1989:Q1
CPI	68.97	57.89	58.93	63.64	59.26	2001:Q4
TBILL	81.03*	92.98***	85.71**	80.00	70.37	2001:Q4

Note: The table reports the empirical coverage rates of 70 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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 80 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. 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, 80 quarter rolling window for constant variance, shorter sample

		SPF horizon					
Variable		0	1	2	3	4	eval. begin
RGDP	(SV rel.)	11.89%	12.57%	12.36%	11.77%	11.53%	1989:Q1
	(CONST abs.)	2.05	2.35	2.47	2.57	2.62	
UNRATE	(SV rel.)	4.22%	5.40%	4.00%	2.58%	0.78%	1989:Q1
	(CONST abs.)	0.15	0.30	0.44	0.57	0.69	
PGDP	(SV rel.)	13.32%	14.49%	16.63%	16.47%	16.48%	1989:Q1
	(CONST abs.)	0.99	1.10	1.15	1.19	1.26	
CPI	(SV rel.)	−13.34%	−11.47%	−11.96%	−18.46%	−14.13%	2001:Q4
	(CONST abs.)	1.27	1.80	1.87	1.79	1.89	
TBILL	(SV rel.)	6.52%	9.62%	7.74%	5.60%	3.24%	2001:Q4
	(CONST abs.)	0.13	0.39	0.62	0.83	1.02	

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 2016:Q1 (evaluated against realized data as far as available). 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 80 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing.

Table 3: 70 percent forecast error coverage rates, 40 quarter rolling window for constant variance, baseline sample

Variable	SPF horizon 0	1	2	3	4	eval. begin
Panel A: SV						
RGDP	67.72	71.43	74.40	72.58	69.11	1984:Q1
UNRATE	75.19	73.44	72.44	73.02	68.00	1984:Q1
PGDP	74.02	73.02	78.40**	78.23**	78.05**	1984:Q1
CPI	70.51	66.23	64.47	65.33	62.16	1996:Q4
TBILL	78.21*	81.82**	72.37	65.33	63.51	1996:Q4
Panel B: CONST						
RGDP	76.38*	76.98*	78.40**	75.00	75.61	1984:Q1
UNRATE	75.19	77.34*	76.38	72.22	72.80	1984:Q1
PGDP	81.10***	81.75***	80.80***	79.84**	78.86**	1984:Q1
CPI	75.64	62.34	60.53	60.00	59.46	1996:Q4
TBILL	74.36	85.71***	76.32	73.33	66.22	1996:Q4

Note: The table reports the empirical coverage rates of 70 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). 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 40 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. 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, 40 quarter rolling window for constant variance, baseline sample

		SPF horizon					
Variable		0	1	2	3	4	eval. begin
RGDP	(SV rel.)	7.93%	8.90%	8.65%	8.12%	7.70%	1984:Q1
	(CONST abs.)	2.08	2.35	2.45	2.53	2.59	
UNRATE	(SV rel.)	3.39%	2.15%	1.90%	0.69%	0.11%	1984:Q1
	(CONST abs.)	0.15	0.30	0.44	0.56	0.68	
PGDP	(SV rel.)	6.40%	6.42%	7.16%	6.58%	6.50%	1984:Q1
	(CONST abs.)	1.00	1.09	1.15	1.18	1.23	
CPI	(SV rel.)	−0.58%	−1.21%	−2.12%	−3.02%	−4.57%	1996:Q4
	(CONST abs.)	1.25	1.74	1.81	1.81	1.81	
TBILL	(SV rel.)	6.65%	8.49%	7.27%	5.72%	4.18%	1996:Q4
	(CONST abs.)	0.12	0.40	0.65	0.87	1.08	

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 2016:Q1 (evaluated against realized data as far as available). 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 40 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing.

Table 5: 70 percent forecast error coverage rates, 20 quarter rolling window for constant variance, baseline sample

	SPF horizon					
Variable	0	1	2	3	4	eval. begin
			Panel A: SV			
RGDP	67.72	71.43	74.40	72.58	69.11	1984:Q1
UNRATE	75.19	73.44	72.44	73.02	68.00	1984:Q1
PGDP	74.02	73.02	78.40**	78.23**	78.05**	1984:Q1
CPI	70.51	66.23	64.47	65.33	62.16	1996:Q4
TBILL	78.21*	81.82**	72.37	65.33	63.51	1996:Q4
			Panel B: CONST			
RGDP	77.17**	74.60	78.40**	77.42	78.05*	1984:Q1
UNRATE	79.07**	77.34*	74.80	76.19	72.80	1984:Q1
PGDP	80.31***	80.95***	82.40***	83.06***	81.30**	1984:Q1
CPI	75.64	64.94	65.79	69.33	64.86	1996:Q4
TBILL	80.77**	87.01***	78.95	77.33	72.97	1996:Q4

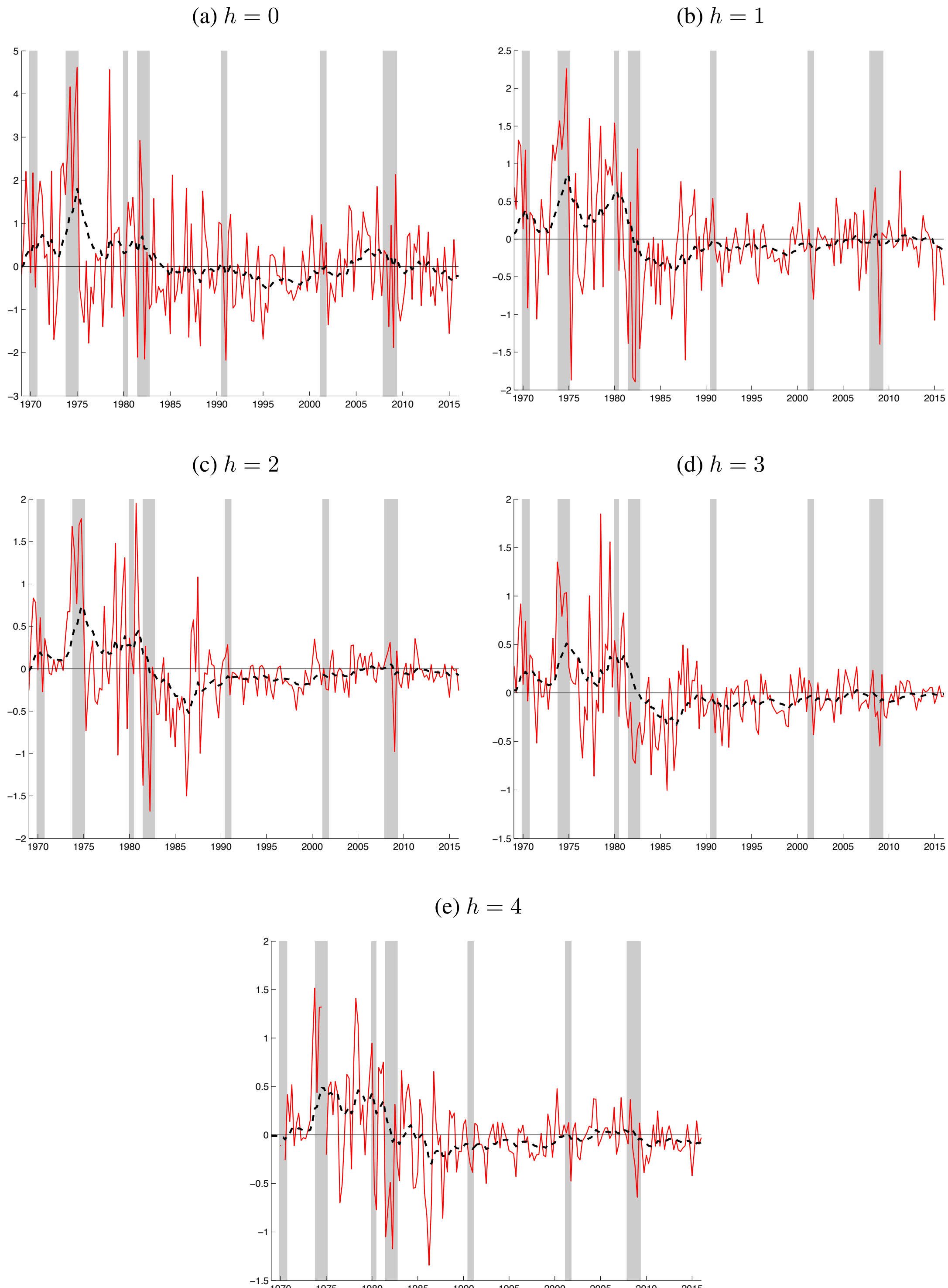
Note: The table reports the empirical coverage rates of 70 percent out-of-sample prediction intervals. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). The upper panel provides results based on our proposed multi-horizon SV model. The lower panel provides results based on a constant-variance model estimated over rolling windows with 20 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing. Statistically significant departures from correct nominal coverage are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 6: Density forecast accuracy as measured by CRPS, 20 quarter rolling window for constant variance, baseline sample

		SPF horizon					
Variable		0	1	2	3	4	eval. begin
RGDP	(SV rel.)	7.79%	9.50%	9.63%	8.92%	8.29%	1984:Q1
	(CONST abs.)	2.08	2.37	2.48	2.56	2.60	
UNRATE	(SV rel.)	7.57%	5.67%	6.40%	5.28%	3.72%	1984:Q1
	(CONST abs.)	0.16	0.32	0.46	0.59	0.71	
PGDP	(SV rel.)	7.40%	7.39%	7.54%	6.76%	6.37%	1984:Q1
	(CONST abs.)	1.01	1.11	1.15	1.18	1.23	
CPI	(SV rel.)	7.86%	5.86%	4.89%	4.24%	−5.05%	1996:Q4
	(CONST abs.)	1.37	1.87	1.94	1.94	1.81	
TBILL	(SV rel.)	10.74%	11.15%	9.78%	8.73%	7.34%	1996:Q4
	(CONST abs.)	0.13	0.42	0.67	0.90	1.12	

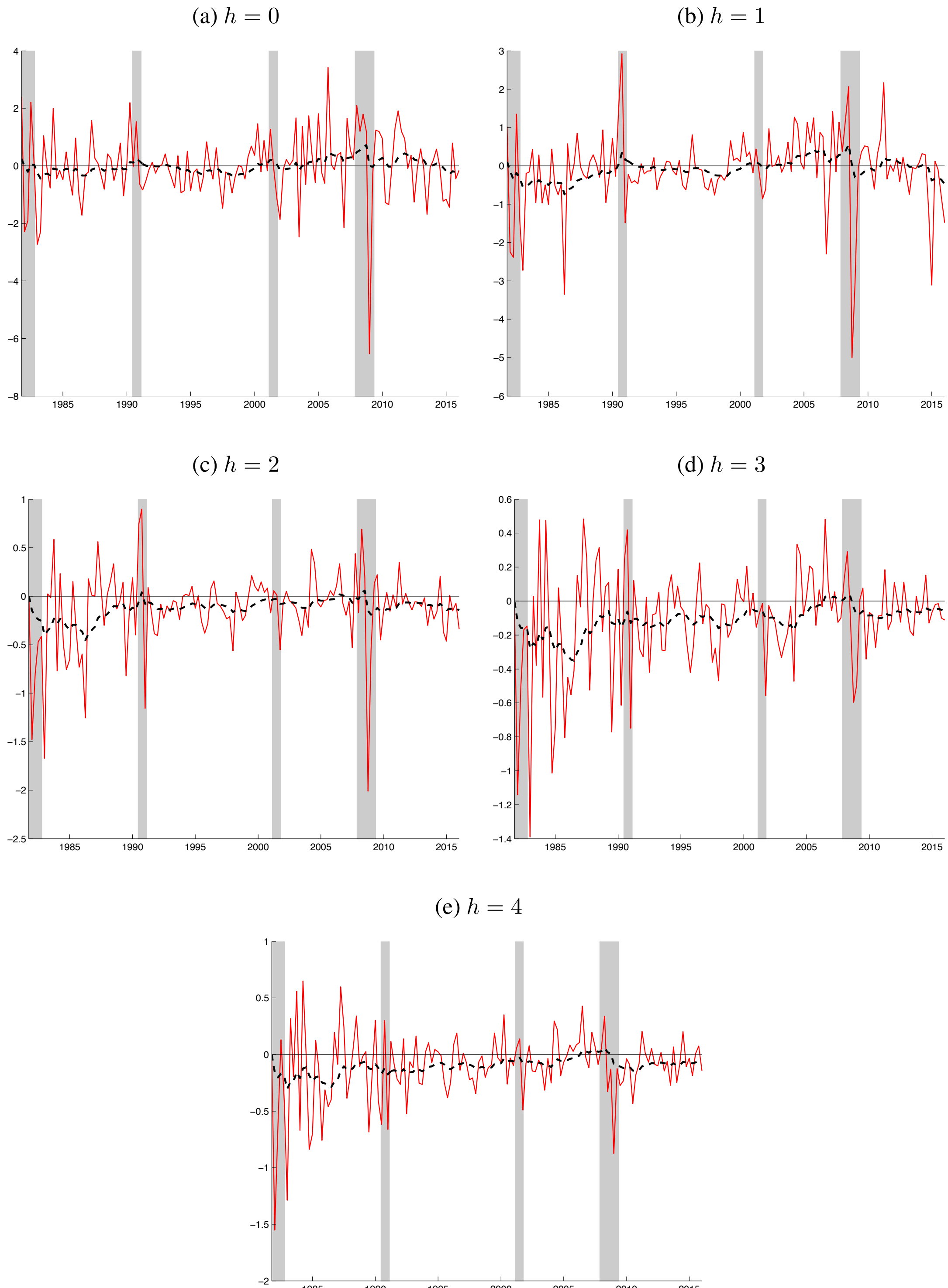
Note: The table reports CRPS results for out-of-sample density forecasts. The sample uses predictions made from the date given in the right-most column through 2016:Q1 (evaluated against realized data as far as available). For each variable, the top row reports the relative CRPS calculated as the percentage 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 20 quarterly observations. Prior to estimation and evaluation, SPF forecast errors have been demeaned on a rolling basis, using means estimated with (one-sided) exponential smoothing.

Figure 1: Expectational Updates for GDP Deflator Inflation



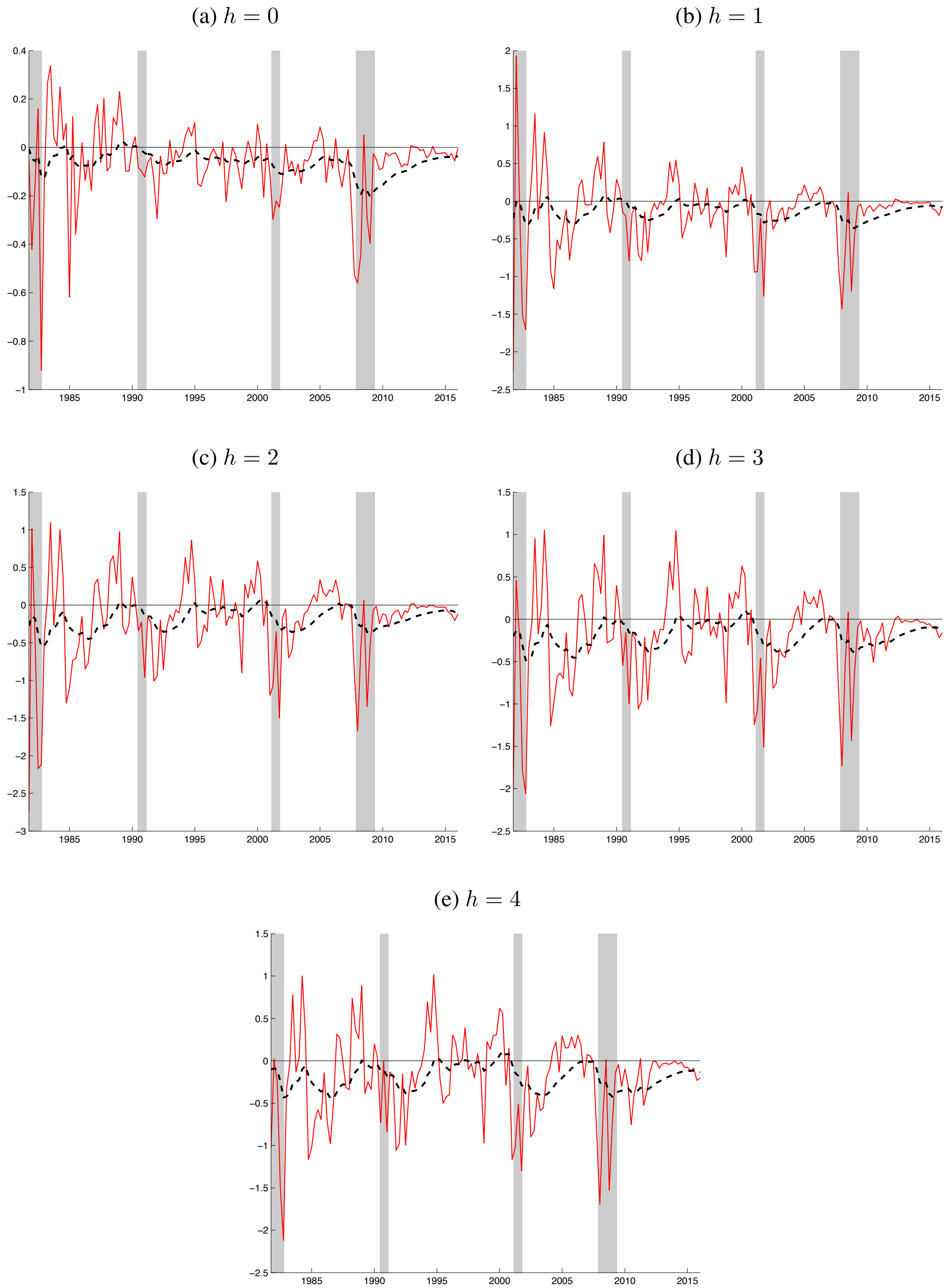
Note: The figure reports (in the red lines) the elements of the vector of expectational updates η_t used in model estimation, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 2: Expectational Updates for CPI Inflation



Note: The figure reports (in the red lines) the elements of the vector of expectational updates η_t used in model estimation, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

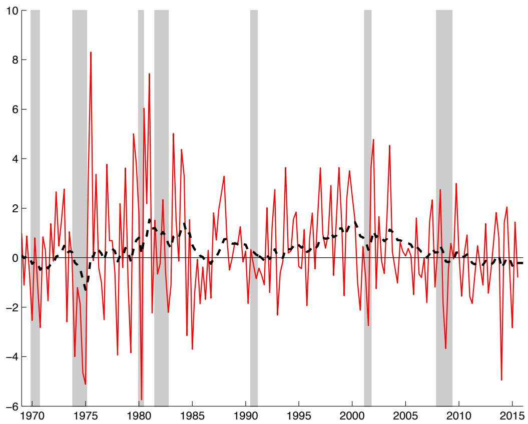
Figure 3: Expectational Updates for the Three-Month Tbill Rate



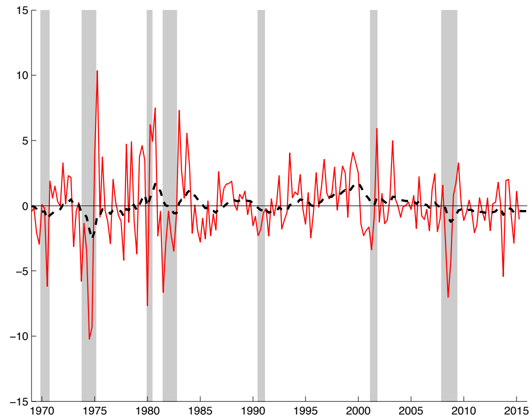
Note: The figure reports (in the red lines) the elements of the vector of expectational updates η_t used in model estimation, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 4: Forecast Errors for Real GDP Growth

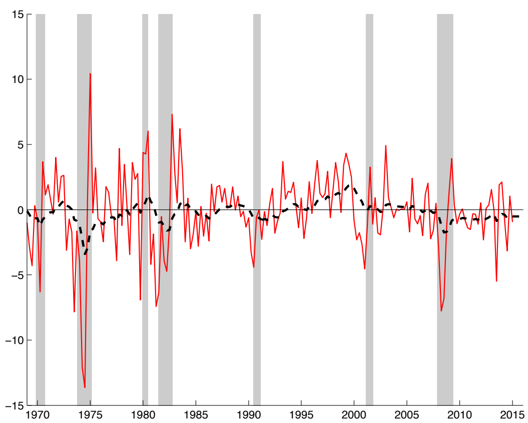
(a) $h = 0$



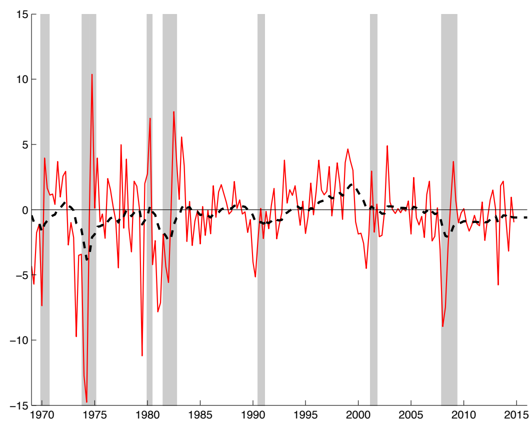
(b) $h = 1$



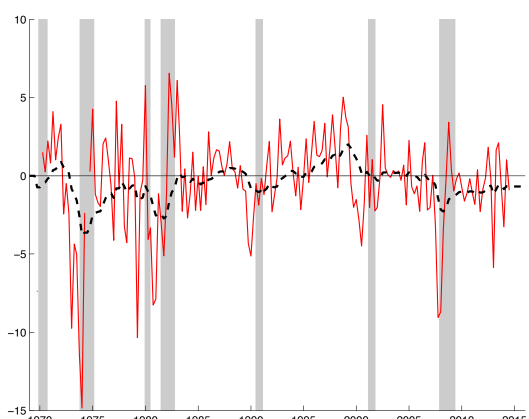
(c) $h = 2$



(d) $h = 3$



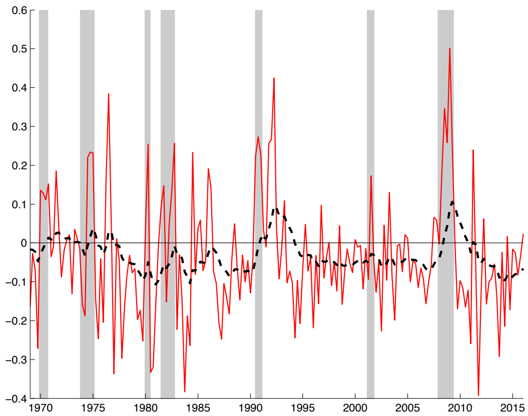
(e) $h = 4$



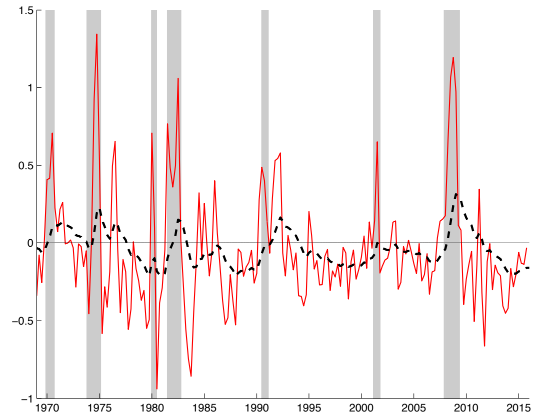
Note: The figure reports (in the red lines) SPF forecast errors, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 5: Forecast Errors for the Unemployment Rate

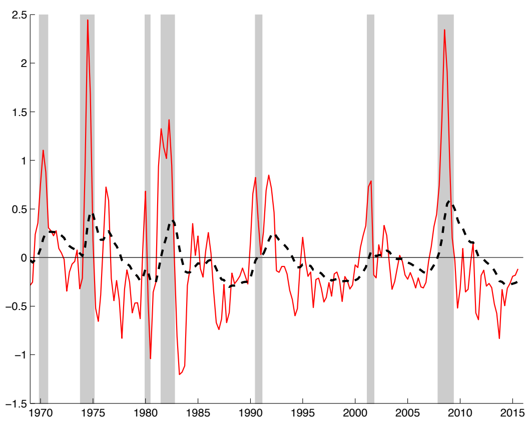
(a) $h = 0$



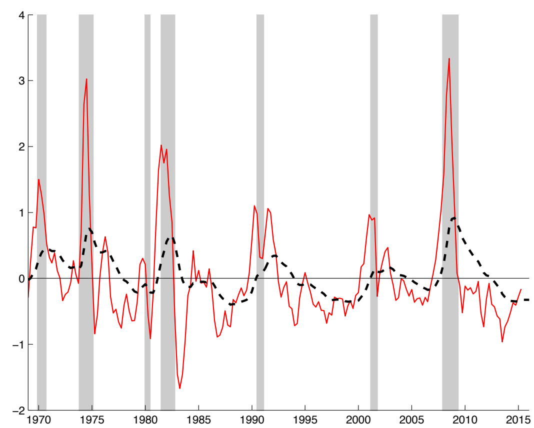
(b) $h = 1$



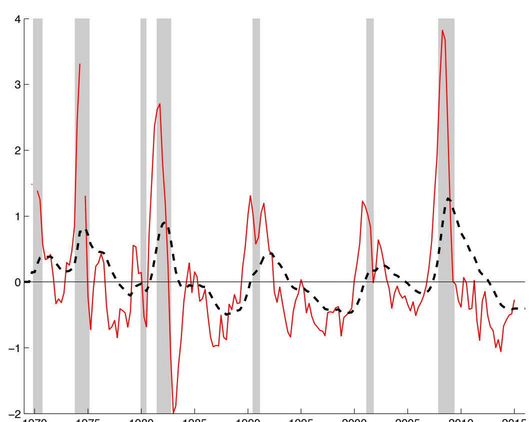
(c) $h = 2$



(d) $h = 3$



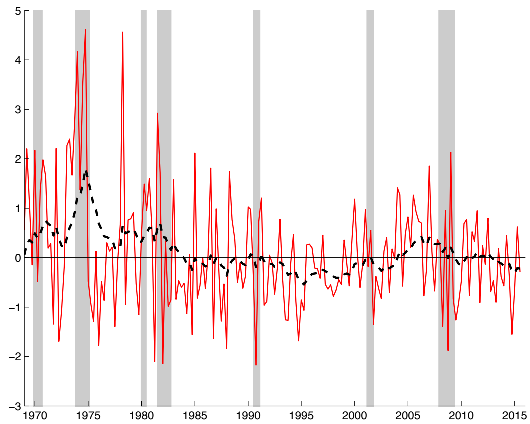
(e) $h = 4$



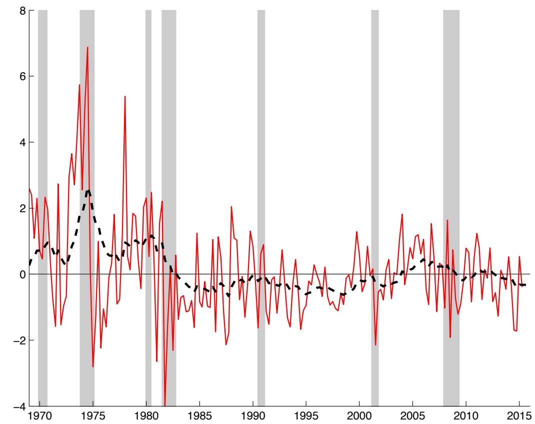
Note: The figure reports (in the red lines) SPF forecast errors, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 6: Forecast Errors for GDP Deflator Inflation

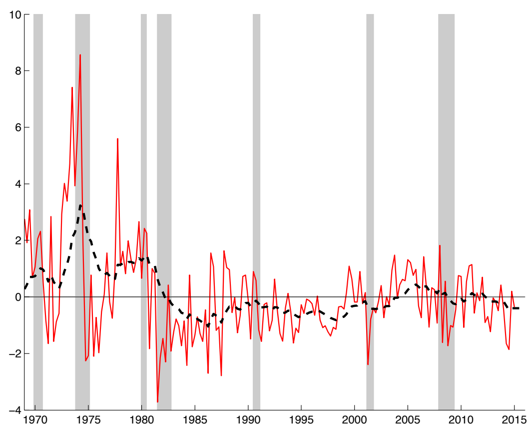
(a) $h = 0$



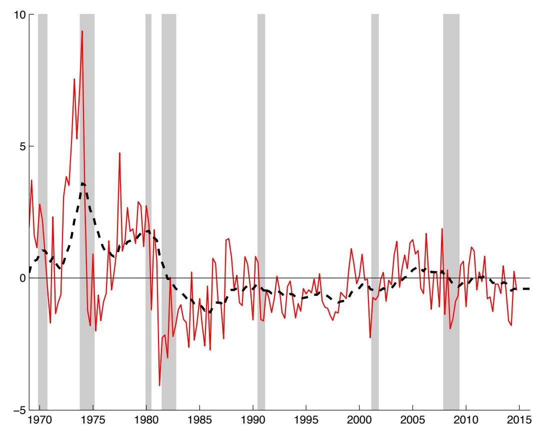
(b) $h = 1$



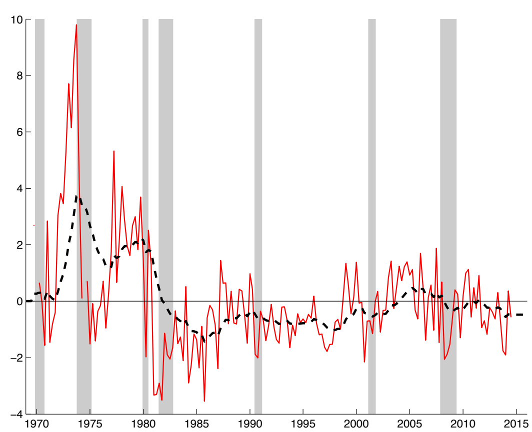
(c) $h = 2$



(d) $h = 3$

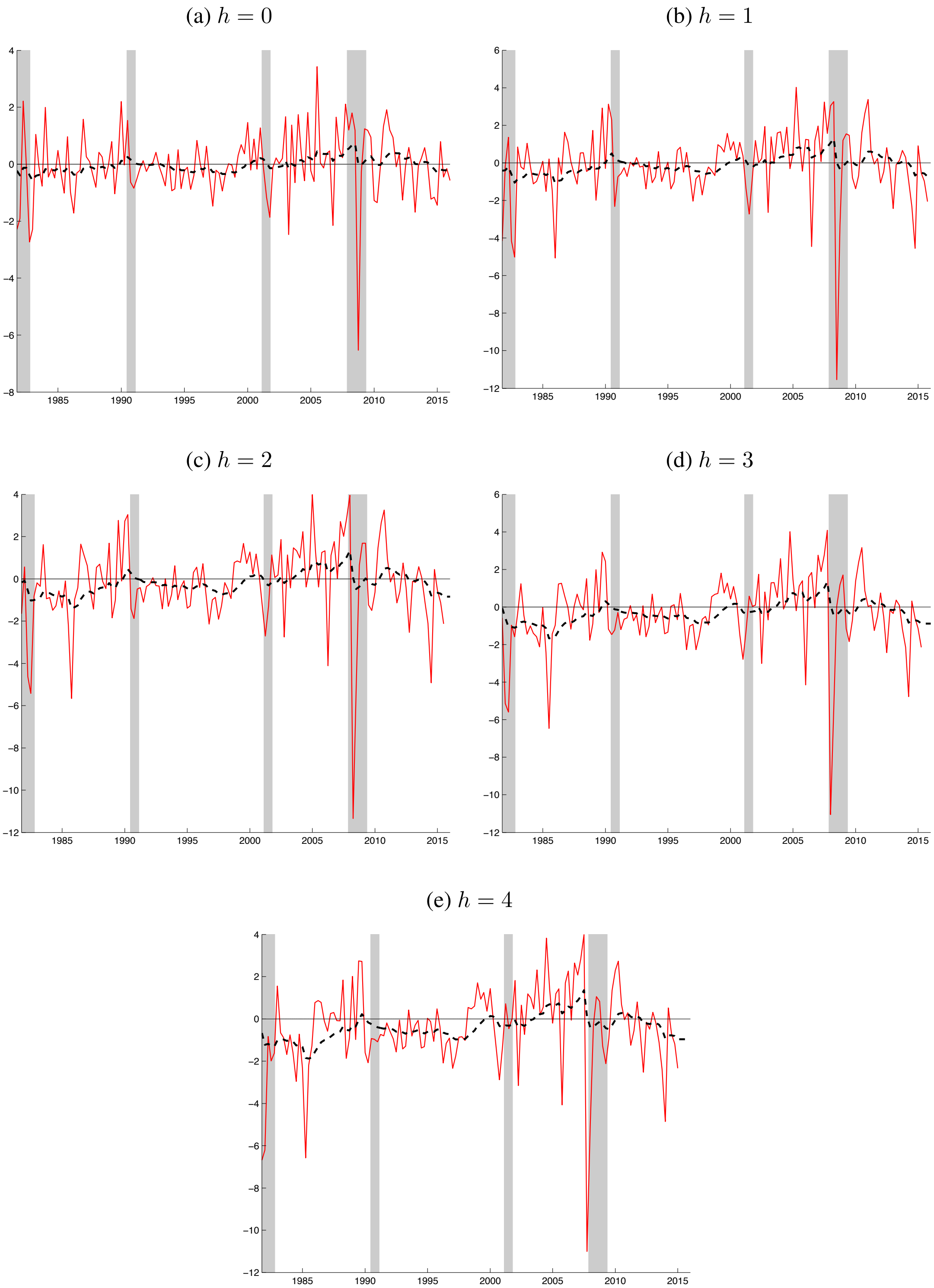


(e) $h = 4$



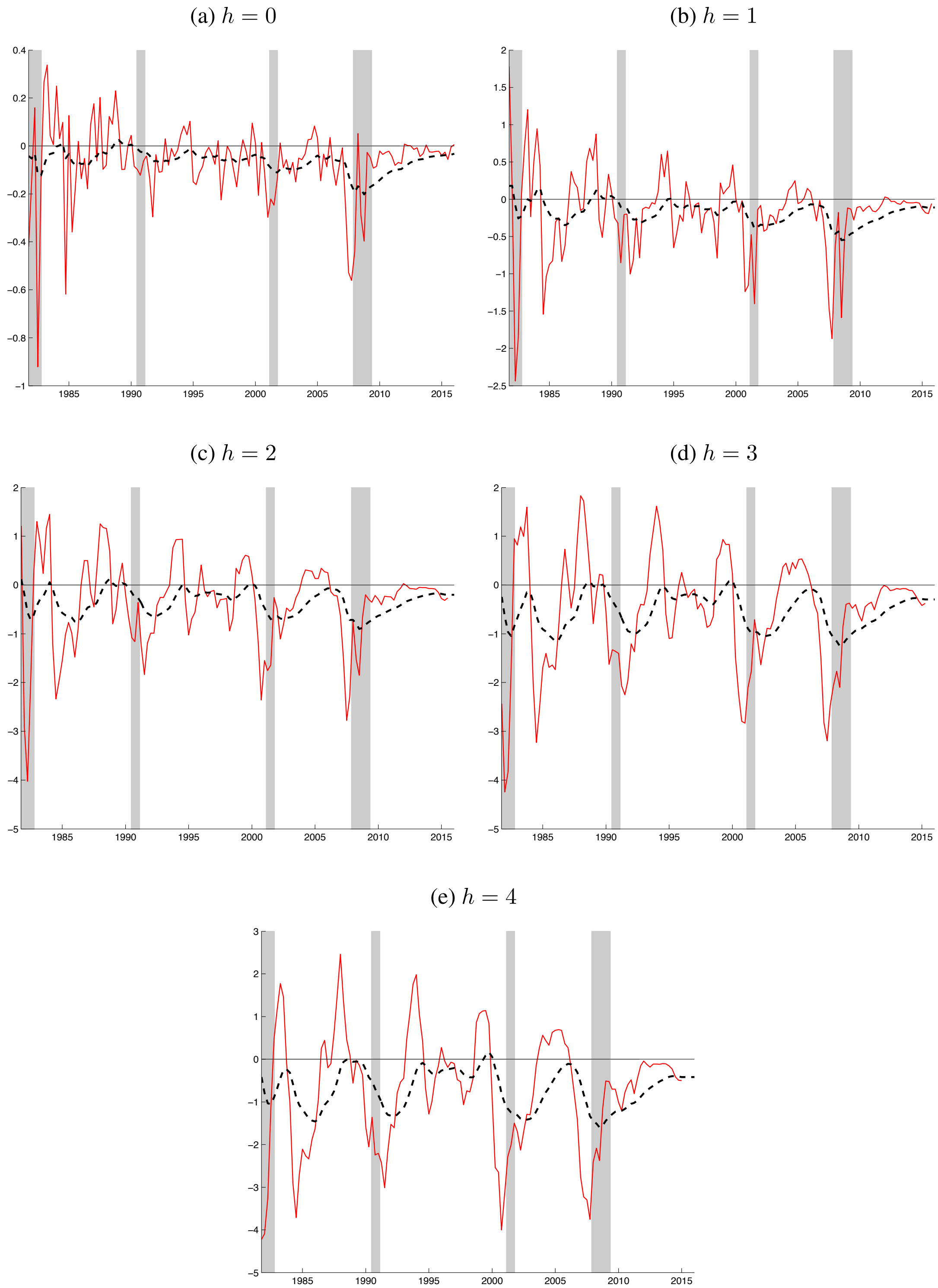
Note: The figure reports (in the red lines) SPF forecast errors, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 7: Forecast Errors for CPI Inflation



Note: The figure reports (in the red lines) SPF forecast errors, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.

Figure 8: Forecast Errors for the Three-Month Tbill Rate



Note: The figure reports (in the red lines) SPF forecast errors, along with (in dotted black lines) the time series of estimates of means obtained with one-sided exponential smoothing. NBER recessions are indicated by gray bars.