

w o r k i n g
p a p e r

17 15R

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

Todd E. Clark, Michael W. McCracken, and
Elmar Mertens



FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working papers of the Federal Reserve Bank of Cleveland are preliminary materials circulated to stimulate discussion and critical comment on research in progress. They may not have been subject to the formal editorial review accorded official Federal Reserve Bank of Cleveland publications. 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, Federal Reserve System, or the Bank for International Settlements.

Working papers are available on the Cleveland Fed's website:
<https://clevelandfed.org/wp>

Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors

Todd E. Clark, Michael W. McCracken, and Elmar Mertens

We estimate uncertainty measures for point forecasts obtained from survey data, pooling information embedded in observed forecast errors for different forecast horizons. To track time-varying uncertainty in the associated forecast errors, we derive a multiple-horizon specification of stochastic volatility. We apply our method to forecasts for various macroeconomic variables from the Survey of Professional Forecasters. Compared to constant variance approaches, our stochastic volatility model improves the accuracy of uncertainty measures for survey forecasts. Our method can also be applied to other surveys like the Blue Chip Consensus, or the Federal Open Market Committee's Summary of Economic Projections.

JEL codes: E37, C53.

Keywords: Stochastic volatility, survey forecasts, fan charts.

Suggested citation: Clark, Todd E., Michael W. McCracken, and Elmar Mertens, 2018. "Modeling Time-Varying Uncertainty of Multiple-Horizon Forecast Errors," Federal Reserve Bank of Cleveland, Working Paper no. 17-15R. <https://doi.org/10.26509/frbc-wp-201715r>.

Todd E. Clark (corresponding author) is at the Federal Reserve Bank of Cleveland (todd.clark@clev.frb.org), Michael W. McCracken is at the Federal Reserve Bank of St. Louis (michael.w.mccracken@stls.frb.org), and Elmar Mertens is at the Bank for International Settlements (elmar.mertens@bis.org). The authors gratefully acknowledge Tom Stark's help with the Philadelphia Fed's real-time data sets and helpful discussions with Malte Knuppel, Serena Ng, Jonathan Wright, and seminar or conference participants at the BIS, Federal Reserve Bank of St. Louis, University of Montreal, University of Pennsylvania, winter 2018 Econometric Society meeting, 2017 SNDE meeting, 2017 IAAE meeting, 2017 NBER Summer Institute, 2017 Bundesbank workshop on forecasting, and the CIRANO/CIREQ/Philadelphia Fed conference on real-time data analysis.

*First version September 2017.

1 Introduction

Macroeconomic forecasts play a key role in the monetary policy communications of central banks. These projections are commonly presented in charts that include point forecasts and estimates of uncertainty around the forecast. A number of central banks use the size of historical forecast errors to quantify forecast uncertainty.¹ For example, the Reserve Bank of Australia and the European Central Bank publish forecast fan charts with uncertainty bands derived from historical forecast errors. The Federal Reserve's Federal Open Market Committee's (FOMC) Summary of Economic Projections (SEP) includes forecast fan charts with uncertainty bands computed with the root mean square errors (RMSEs) of historical forecasts. The historical RMSEs are intended to provide an approximate 70 percent confidence interval around the forecast indicated by the median of the FOMC participants' projections.²

In the estimates used in such charts, central banks commonly treat uncertainty as constant over time, with a sample period chosen judiciously. As examples, the historical accuracy measures published in the Federal Reserve's SEP are based on a 20-year window of forecast errors, and the fan charts of the Bank of England are constructed using information that includes measures of forecast accuracy over the previous 10 years. The Reserve Bank of Australia estimates uncertainty with forecast errors since 1993. A rolling window of forecast errors or a particular sample start can be seen as a simple approach to accommodating some change in forecast uncertainty over time. For example, structural changes such as the Great Moderation or unusual periods such as the recent

¹The supplemental appendix provides links to documents for the countries we mention: the Reserve Bank of Australia, the European Central Bank, the Federal Reserve, and the Bank of England. Knuppel and Schulteffer (2012) and Tulip and Wallace (2012) summarize the approaches used by a broader range of central banks.

²As detailed in Reifschneider and Tulip (2007, 2017), the RMSEs are computed from the errors of several different forecasts, including, among others, the Survey of Professional Forecasters, the Blue Chip Consensus, and the Congressional Budget Office.

Great Recession can lead to significant shifts in the sizes of forecast errors and, in turn, forecast uncertainty. Reifschneider and Tulip (2017) provide simple evidence of changes in the sizes of forecast errors associated with projections from the Federal Reserve and other sources, including the Survey of Professional Forecasters (SPF) and Blue Chip Consensus. Possible time variation in forecast error variances is important because failing to capture it may result in forecast confidence bands that are either too wide or too narrow.

Although the approaches commonly used at central banks make some allowance for possible shifts in forecast uncertainty over time, a fairly large literature on the forecast performance of time series and structural economic models suggests it may be possible to improve estimates of forecast uncertainty by more explicitly modeling time variation in forecast error variances.³ In this literature, time variation in the size of estimated forecast errors turns out to be large, and modeling it significantly improves the accuracy and calibration of density forecasts. Most such studies have focused on vector autoregressive (VAR) models with one particular formulation of time-varying volatility in forecast errors: stochastic volatility. Examples include Carriero, Clark, and Marcellino (2016), Clark (2011), Clark and Ravazzolo (2015), and D'Agostino, Gambetti, and Giannone (2013). Diebold, Schorfheide, and Shin (2017) provide similar evidence for dynamic stochastic general equilibrium (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 central banks and professional forecasters such as the SPF 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 typical models. In parametric time series models, one-step-ahead predictions and errors provide the basis of the specification and estimation; multi-step

³The forecasting literature builds on the initial work of Cogley and Sargent (2005) and Primiceri (2005) on VARs with stochastic volatility and Justiniano and Primiceri (2008) on DSGE models with stochastic volatility.

errors result from recursion over the sequence of one-step-ahead forecasts generated by the model and do not directly play a role in estimation. But historical errors of sources such as the SPF or Federal Reserve span multiple forecast horizons, with some correlation or overlap across the horizons. For example, at a two-step-ahead forecast horizon, the forecast error for period $t + 2$ will share a component with the two-step error for period $t + 1$, creating serial correlation, and the two-step errors will have some correlation with one-step-ahead errors. No model exists for such correlations in the case in which the multi-step forecast errors (covering multiple horizons) are primitives.

Accordingly, in this paper, we develop a multiple-horizon specification of stochastic volatility for forecast errors from sources such as the Federal Reserve, the SPF, or the Blue Chip Consensus, 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 their width; the explicit modeling of the time variation of volatility eliminates the need for somewhat arbitrary judgments of sample stability. At each forecast origin, we observe the forecast error from the previous quarter and forecasts for the current quarter and several subsequent quarters. To address the challenge of overlap in forecast errors across horizons, we formulate the model to make use of the forecast error from the previous quarter (period $t - 1$) 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 the assumption that the forecasts represent conditional expectations. For this vector of observations, we specify a multiple-horizon stochastic volatility model that can be estimated with Bayesian methods. From the estimates, we are able to compute the time-varying conditional variance of forecast errors at each horizon of interest. Of course, forecasts from sources such as the SPF may not be optimal, such that forecast updates are not entirely serially uncorrelated. As we detail below, we also consider a version of our model extended to allow a low-order VAR specification of the data vector containing forecast updates.

After developing the model and estimation algorithm, we provide a range of results for forecasts of GDP growth, unemployment, inflation, and a short-term interest rate from the SPF, which provides the longest available history of data and has accuracy very similar to Federal Reserve forecasts (Reifschneider and Tulip 2007, 2017). First, we document considerable time variation in historical forecast error variances by estimating the model over the full sample of data for each variable. 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 recent Great Recession. Error variances move together strongly — but not perfectly — across forecast horizons. Second, we produce 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 sample variances computed with rolling windows of forecast errors as in approaches such as Reifschneider and Tulip (2007, 2017).

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

attempt to allow for possible bias and serial correlation in the expectational updates.

The paper proceeds as follows. Section 2 describes the SPF forecasts and data used in the 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. A supplemental appendix provides additional materials, including results based on forecasts from the Federal Reserve’s Greenbook.

2 Data

Reflecting in part the survey forecasts available, we focus on quarterly forecasts for a basic set of major macroeconomic aggregates: GDP growth (RGDP), the unemployment rate (UNRATE), inflation in the GDP price index (PGDP) and CPI, and the 3-month Treasury bill (TBILL, or T-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 the paper’s results on quarterly forecasts from the SPF because they are widely studied and publicly available, and offer the longest available quarterly time series of forecasts. Alternatives such as the Blue Chip Consensus are not available publicly or for as long a sample.

We obtained the SPF forecasts of growth, unemployment, inflation, and the T-bill rate from the Federal Reserve Bank of Philadelphia’s Real-Time Data Set for Macroeconomists (RTDSM). Reflecting the data available, our estimation samples start with 1969:Q1 for GDP growth, unem-

⁴The 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.

ployment, and GDP inflation and 1981:Q4 for CPI inflation and the T-bill rate; the sample end point is 2018:Q1 for forecasts, and 2017:Q4 for readings of realized values. At each forecast origin, the available forecasts typically span five quarterly horizons, from the current quarter through the next four quarters. We form the point forecasts using the mean SPF responses.

To evaluate the forecasts and our model, we also need measures of the outcomes of the variables. In the case of GDP growth and GDP inflation, data can be substantially revised over time. To form confidence bands around the forecast at the time the forecast is produced, in roughly the middle of quarter t , we measure the quarter $t - 1$ forecast error with the first (in time) estimate of the outcome. Specifically, for GDP growth and GDP inflation, we obtain real-time measures for quarter $t - 1$ data as these data were publicly available in quarter t from the quarterly files of real-time data in the 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. We also use the first-available estimate from the RTDSM to measure the outcomes needed to evaluate the forecasts and our models.

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 for these variables. We obtained data on the unemployment rate, CPI, and 3-month T-bill rate from the FRED database of the Federal Reserve Bank of St. Louis.

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, and in principle, it would be interesting to compare our measures against those of the SPF. However, in the SPF, these probability distributions are provided only for fixed-event forecasts (for the current and next calendar year) rather than fixed-horizon forecasts, making it difficult to use the information to compute uncertainty around the fixed-horizon point forecasts of the SPF. Thus, making use of the SPF's probability distributions to compare to our main results is hardly feasible (without some very tenuous assumptions necessary to approximate fixed-horizon

forecasts from fixed-event forecasts). Moreover, some research has documented 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 uncertainty at shorter forecast horizons (Clements 2014).⁵ Clements (2016) finds that density forecasts obtained from the SPF histograms are no more accurate than density forecasts estimated from the historical distributions of past point forecast errors.

Before we turn from the data to our model, note that, as a general matter, our model can be readily applied to forecasts from other sources. As the introduction notes, the forecasts need to be of the fixed-horizon type (not fixed-event) and cover (in sequence) multiple forecast horizons. The forecasts can be at any data frequency, although quarterly would be most typical in macroeconomic settings. Although our data on growth and inflation are quarter-on-quarter percent changes, our model could be applied to use year-on-year percent changes.⁶

3 Model

In this section, we first detail the forecast error decomposition that underlies our proposed model and then present the model, along with an extended version of it. We conclude by describing a constant variance benchmark included in the empirical analysis.

⁵Using 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. In contrast, in an earlier analysis of data from the U.S. SPF, Giordani and Soderlind (2003) find that some GARCH models imply uncertainty estimates that are correlated with *ex ante* uncertainty obtained from probability distribution forecasts.

⁶In this case, the primary changes would relate to the specifics of the aggregation matrix polynomial $\mathbf{B}(L)$ described below.

3.1 Forecast error decomposition

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

In practice, exactly how the forecast is constructed is unknown, except that the forecast likely includes some subjective judgment and need 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 $E_t y_t, E_t y_{t+1}, \dots, E_t y_{t+H}$, 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. Knoppel (2014) develops an approach for estimating forecast accuracy that accounts for such overlap in observed forecast errors, but under the implicit assumption that forecast error variances are constant over time. To model time variation in forecast uncertainty in overlapping forecasts, we make use of a decomposition of the multi-step forecast error into a nowcast error and the sum of changes (from the previous period to the current period) in forecasts for subsequent periods. For our baseline model, we appeal to the martingale difference property of optimal forecasts and treat the vector of forecast updates as serially uncorrelated. However, even without that assumption, our use of this decomposition can be seen as a form of pre-whitening of the multi-step forecast errors, which will be useful for specification of an extended model described further below.

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 and the next h periods for the expected value at $t+h$:

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

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

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

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

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

$$E_{t-1} \mu_{t+h|t} = 0. \quad (2)$$

Assuming that, at every forecast origin t , the forecast source 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.

Under the martingale difference assumption on the expectational updates, we can characterize the conditional variance of the multi-step forecast error as the sum of the (expected) variances of the individual terms in (1):

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

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

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

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 detailed below, this decomposition could be used to build up estimates of $\text{var}_t(e_{t+h})$ from estimates of the conditional variances, for (1) the variance of the nowcast error, $\text{var}_t(e_t)$, and (2) the variance of the expectational update of forecasts for horizon $i = 1, \dots, h$, $\text{var}_t(\mu_{t+i|t+1})$. 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 Model of time-varying volatility

Based on the decomposition (1) and the martingale difference assumption (2), we specify a multivariate stochastic volatility model for the available nowcast error and expectational updates. The supplemental appendix shows that this model is conceptually consistent with a general class of linear forecasting models.

Starting with the data of the model, as noted above the forecast origin is roughly the middle of quarter t , corresponding to the publication of the survey forecast. At the time the forecasters construct their projections, they have data on quarter $t - 1$ and some macroeconomic data on quarter t . We construct a data vector strictly contained in that information set, containing $H + 1$ elements: the nowcast error for quarter $t - 1$ and the revisions in forecasts for outcomes in quarters t through $t + H - 1$.⁷ With the SPF forecasts, $H = 4$: we have the nowcast error and four forecast updates to use.

More specifically, we define the data vector as:

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

This specification includes an offset in timing between the first element of $\boldsymbol{\eta}_t$ and the remaining

⁷Although at origin t the forecasts go through period $t + H$, the available forecast revisions only go through period $t + H - 1$.

elements, by pairing the $t - 1$ nowcast error — the most recently observed nowcast error at the forecast origin t — with the period 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. Based on the accounting identity (1), given the vector $\boldsymbol{\eta}_t$, we are able to obtain the forecast errors from:

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

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

We build the baseline specification around an assumption that the data vector $\boldsymbol{\eta}_t$ of expectational updates and the forecast errors \mathbf{e}_t have means of zero. Reifschneider and Tulip (2017) also assume future forecasts to be unbiased, treating any past historical bias as transitory. In our case, in preliminary analysis, we obtained similar results when, before estimating the model, we demeaned the elements of the data vector $\boldsymbol{\eta}_t$ using a real-time approach to computing a time-varying mean, with one-sided exponential smoothing. The next subsection presents a generalization that allows constant non-zero means.

Our baseline model of the expectational updates is a multivariate stochastic volatility specification, allowing for correlation both across elements of $\boldsymbol{\eta}_t$ as well as across innovations to the log volatilities of each component of $\boldsymbol{\eta}_t$ (i.e., between both level and scale shocks to $\boldsymbol{\eta}_t$):

$$\begin{aligned} \boldsymbol{\eta}_t &= \mathbf{A}\tilde{\boldsymbol{\eta}}_t & \mathbf{A} &= \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ a_{H+1,1} & a_{H+1,2} & \dots & & 1 \end{bmatrix} \\ \tilde{\boldsymbol{\eta}}_t &= \boldsymbol{\Lambda}_t^{0.5} \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t \sim i.i.d. N(\mathbf{0}, \mathbf{I}_{H+1}), \boldsymbol{\Lambda}_t \equiv \text{diag}(\lambda_{1,t}, \dots, \lambda_{H+1,t}) \\ \log(\lambda_{i,t}) &= \log(\lambda_{i,t-1}) + \nu_{i,t}, \quad i = 1, \dots, H + 1 \\ \boldsymbol{\nu}_t &\equiv (\nu_{1,t}, \nu_{2,t}, \dots, \nu_{H+1,t})' \sim i.i.d. N(\mathbf{0}, \boldsymbol{\Phi}), \end{aligned} \quad (7)$$

where \mathbf{A} , a lower triangular matrix with values of 1 on the diagonal, serves to capture correlations across the components of $\boldsymbol{\eta}_t$, while $\boldsymbol{\Phi}$ captures correlations across the innovations to stochastic volatility. 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 $\boldsymbol{\eta}_t$ 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. For such 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).⁸ Section 5's robustness section summarizes results for a version of the model with volatilities restricted to follow a single common factor process.

By choosing an otherwise conventional, conditionally linear and Gaussian data-generating process, our approach will yield prediction intervals and densities that are symmetric. In doing so, we follow the broader literature (see references above) on including stochastic volatility in time series models for macroeconomic forecasting. Although the model makes use of conditional innovations (in $\boldsymbol{\varepsilon}_t$) that are Gaussian, this does not imply that the observed forecast errors and expectational updates are Gaussian. In fact, the model implies that the distributions of the observed expectational updates and forecast errors feature fat tails. We leave as a subject for further research the extension of the model to allow fat tails in the conditional errors $\boldsymbol{\varepsilon}_t$, drawing on the specification of Jacquier, Polson, and Rossi (2004) or the 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, and Pinter 2017; Clark and Ravazzolo 2015;

⁸We obtained similar results for a model treating the volatility innovations as mutually independent (as in, e.g., Cogley and Sargent 2005).

Curdia, Del Negro, and Greenwald 2015). Stock and Watson (2016) find that a related mixture-of-normals approach to filtering inflation outliers is helpful.⁹

3.3 Generalized model without MDS assumption

As noted above, our baseline specification reflects an assumption that the vector of expectational updates forms a martingale difference sequence, consistent with full rationality of the forecasts. This assumption helps to yield a parsimonious model, and parsimony is well known to be helpful in forecasting. However, studies such as Croushore (2010) and Reifschneider and Tulip (2017) provide evidence of some biases in forecasts from sources such as the SPF and the Greenbook. Moreover, recent research by Coibion and Gorodnichenko (2015) and Mertens and Nason (2018), among others, has shown that survey-based forecasts display information rigidities, reflected in some serial correlation in forecast errors.

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

$$\boldsymbol{\eta}_t = \mathbf{C}_0 + \mathbf{C}_1 \boldsymbol{\eta}_{t-1} + \mathbf{A} \boldsymbol{\Lambda}_t^{0.5} \boldsymbol{\varepsilon}_t, \quad (8)$$

where \mathbf{C}_0 is a vector of intercepts, \mathbf{C}_1 is a matrix of slope coefficients, and the remainder of the model is defined as in the baseline specification. Although the model could easily be extended to include longer lags, we have deliberately chosen to include just one lag, for parsimony and the

⁹We actually examined extending our model to include the Stock-Watson mixture. In our setting with the SPF forecasts, this model extension helped to reduce the influence of some outliers on our stochastic volatility estimates but had little effect on our baseline forecast results.

likely low-order serial correlation in the expectational updates. Although deviations from forecast rationality might induce serial correlation in $\boldsymbol{\eta}_t$, the transformation from forecast errors to forecast updates still serves as a pre-whitening step, given that deviations from rationality in survey forecasts appear limited. As a simple check of the serial correlation in the expectational updates, for each variable we estimated vector autoregressions based on the vector $\boldsymbol{\eta}_t$ using 0 to 4 lags and assessed their fit with the BIC. The BIC indicates that the optimal lag order is 0 for the unemployment rate and CPI inflation and 1 for GDP growth, GDP inflation, and the T-bill rate. Our use of a VAR with one lag appears to be consistent with this simple check.

As detailed below, we estimate this extended model — referred to below as the VAR-SV specification — with conventional Minnesota-type priors on \mathbf{C}_0 and \mathbf{C}_1 .¹⁰ As with the baseline model, we obtain the forecast errors using the accounting identity $\mathbf{e}_t = \mathbf{B}(L)\boldsymbol{\eta}_{t+1}$.

3.4 Estimating the model and forecast uncertainty

The baseline model of (7) and the extension (8) can be estimated by Bayesian Markov Chain Monte Carlo (MCMC) methods. We focus on describing the estimation of the baseline specification; the estimation of the VAR model involves adding a conventional Gibbs step to draw the VAR coefficients from their conditional posterior (see, e.g., Clark and Ravazzolo 2015). The baseline model’s algorithm involves iterating over 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 Cholesky decomposition of $\boldsymbol{\eta}_t$ into $\tilde{\boldsymbol{\eta}}_t$. Second, we estimate the stochastic volatilities of $\tilde{\boldsymbol{\eta}}_t$ using the multivariate version of the Kim, Shephard, and Chib (1998) [henceforth, KSC] algorithm introduced into macroeconomics by Primiceri (2005), and as refined by Omori, et al. (2007). Third, given draws for the sequences of $\log(\lambda_{i,t})$ for all i

¹⁰For the VAR’s coefficients, the prior means are all zero, and the standard deviations take the Minnesota form, with the hyperparameter governing overall shrinkage set at 0.2, the hyperparameter for “other” lags relative to “own” lags set at 0.5, and the hyperparameter governing intercept shrinkage set at 1.

and t we estimate the variance-covariance matrix of innovations to the SV processes, Φ , using an inverse Wishart prior centered around a mean equal to a diagonal matrix with 0.2^2 on its diagonal using $9 + H$ degrees of freedom, which makes the prior slightly informative. 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).

As regards the volatility step, let $\tilde{\eta}_{i,t}$ refer to the i -th element of $\tilde{\boldsymbol{\eta}}_t$. Taking the log of $\tilde{\eta}_{i,t}^2$ yields the measurement equation

$$\log \tilde{\eta}_{i,t}^2 = \log \lambda_{i,t} + \log \varepsilon_{i,t}^2, \quad i = 1, \dots, H + 1, \quad (9)$$

with corresponding transition equation

$$\log(\lambda_{i,t}) = \log(\lambda_{i,t-1}) + \nu_{i,t}, \quad i = 1, \dots, H + 1. \quad (10)$$

As the different elements of $\tilde{\boldsymbol{\eta}}_t$ (as opposed to $\boldsymbol{\eta}_t$) are mutually uncorrelated, the measurement equation (9) includes a χ^2 -distributed innovation that is independent across i . KSC and Omori, et al. (2007) develop mixture-of-normals approximations 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 the SPF forecasts in our sample. KSC, Primiceri (2005), and Del Negro and Primiceri (2015) provide additional detail.

To estimate the uncertainty around multi-step forecasts, we simulate the posterior distribution of forecast errors using the model in (7). 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). For each draw of parameters of the MCMC algorithm, we obtain draws of these terms by simulating forward the vector $\boldsymbol{\eta}_t$ of our multivariate SV model, to obtain, via equation (6), the posterior distribution of forecast errors, via the following steps:

1. For each component i of $\tilde{\boldsymbol{\eta}}_t$, 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 Φ .

2. Simulate the time path of $N(\mathbf{0}, \mathbf{I}_{H+1})$ innovations $\boldsymbol{\varepsilon}_t$ forward from period $t + 1$ through period $t + H + 1$.
3. Obtain the time path of $\tilde{\boldsymbol{\eta}}_{t+h}$ from period $t + 1$ through period $t + H + 1$ as the product of the simulated $\Lambda_{t+h}^{0.5}$ and $\boldsymbol{\varepsilon}_{t+h}$.
4. Transform $\tilde{\boldsymbol{\eta}}_t$ 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).

Given the set of draws produced by this algorithm, we compute the forecast statistics of interest. For example, we compute the standard deviation of the forecast errors and the percentage of observations falling within a plus/minus one-standard-deviation band.

3.5 An alternative approach assuming constant variances of forecast errors

In light of common central bank practice (e.g., Reifschneider and Tulip 2007, 2017 and the fan charts in the Federal Reserve’s SEP), the most natural benchmark against which to compare our proposed model-based approach is one based on historical forecast error variances treated as constant over some window of time and into the future. That is, at each forecast origin t , prediction intervals and forecast densities can be computed assuming normally distributed forecast errors with variance equal to the variance of historical forecast errors over the most recent R periods.¹¹ Accordingly, we report results obtained under such an approach, where we collect continuously updated estimates generated from rolling windows of forecast errors covering the most recent $R =$

¹¹A number of other studies, such as Kenny, Kostka, and Masera (2014) and Rossi and Sekhposyan (2014), have also used normal distributions based on a given point forecast and error variance.

60 quarterly observations. For simplicity, below we will refer to this specification as the “constant variance” approach and denote it with “FE-CONST,” even though it acknowledges the potential for variance changes over time by using a rolling window of observations. Note, too, that this benchmark approach differs from our model-based approach in that the benchmark uses forecast errors directly, whereas our approach uses expectational updates and obtains forecast errors as linear combinations of the updates. In addition, the FE-CONST approach differs in that it relies merely on sample moments without specifying an explicit probability model for the data. Section 5’s robustness section summarizes results for a more parametric rolling window approach based on the vector of expectational updates η_t and an assumed normal distribution for the updates.

Of course, a key choice is the size of the rolling window (R) used in the constant variance approach. As noted above, some central banks use windows of 40 or 80 quarterly observations; Clements (2016) uses 50 quarterly observations. In our analysis, there is an important sample tradeoff in data availability: making the rolling window bigger shortens the forecast sample available for evaluation. Accordingly, in our baseline results, we essentially split the difference, so to speak, and use a rolling window of 60 observations in the constant variance benchmark. With this setting, we have available the following samples for the evaluation of the SPF forecasts: 1984:Q1–2018:Q1 for GDP growth, unemployment, and GDP inflation; and 1996:Q4–2018: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.

4 Evaluation Metrics

The previous section described three alternative volatility models: our proposed stochastic volatility model, our extension to a VAR with stochastic volatility, and a constant variance benchmark. This section describes 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 T-bill rate fall inside one-standard-deviation prediction intervals. We compare these coverage rates to the nominal coverage rate implied by the percentiles of the normal distribution for the area between plus/minus a one-standard-deviation error; up to rounding this covers 68 percent. (We focus on one-standard-deviation/68 percent coverage rates because there are far fewer observations available for evaluating accuracy further out in the tails of the distributions.) A frequency of more (less) than 68 percent means that, on average over a given sample, the estimated forecast density is too wide (narrow). We judge the significance of the results using p -values of t -statistics for the null hypothesis that the empirical coverage rate equals the nominal rate of 68 percent; we compute the t -statistics with the HAC-robust variance estimate of Newey and West (1987) and a lag order equal to the SPF forecast horizon plus 2.¹²

Our second measure of density accuracy is the continuous ranked probability score (CRPS). Some researchers (e.g., Gneiting and Raftery 2007 and Gneiting and Ranjan 2011) prefer the CRPS over the log score because it does a better job of rewarding values from the predictive density that are close to but not equal to the outcome, and it is less sensitive to outlier outcomes. The CRPS, defined such that a lower number is a better score, is given by

$$CRPS_t(y_{t+h}^o) = \int_{-\infty}^{\infty} (F(z) - 1\{y_{t+h}^o \leq z\})^2 dz = E_f |Y_{t+h} - y_{t+h}^o| - 0.5 E_f |Y_{t+h} - Y'_{t+h}|, \quad (11)$$

¹²We also verified the robustness of the significance results to alternative lag-order choices for the Newey-West estimator; specifically, we also used a substantially wider window equal to 2 times the SPF forecast horizon and the automatic lag-order suggestion from Newey and West (1994), $4(N/100)^{2/9}$, where N is the number of observations in the evaluation window. The reported significance levels are generally robust using either lag-order choice.

where F denotes the cumulative distribution function associated with the predictive density f , $1\{y_{t+h}^o \leq z\}$ denotes an indicator function taking value 1 if the outcome $y_{t+h}^o \leq z$ and 0 otherwise, and Y_{t+h} and Y'_{t+h} are independent random draws from the posterior predictive density. We compute the CRPS using the empirical CDF-based approximation given in equation (9) of Krueger, et al. (2017). We gauge the significance of differences in CRPS on the basis of p -values of t -statistics for equality of average CRPS, using HAC-robust variances computed with the Newey and West (1987) estimator and a lag order equal to the SPF forecast horizon plus 2.¹³

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. 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, we take the point forecasts as given from the SPF, whereas, in these papers, the point forecasts vary with each model. 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 model's estimated parameters and in turn its point forecasts. As a result, the evidence on density forecast accuracy from the VAR and DSGE literature commingles the effects of conditional means and variances with rolling windows versus other estimators and other models. In this paper, by using point forecasts from the SPF, we are isolating influences on density accuracy due to variances.

5 Results

We begin this section of results 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 remainder of the section discusses some robustness checks, including results for the VAR-SV extension.

¹³Our results are generally robust to using any of the alternative lag-order choices described above in the context of evaluating coverage rates; see Footnote 12.

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 η_t . In the interest of brevity, here we briefly describe some notable features of the data; figures displaying the data — in the form of both expectational updates and forecast errors — are shown in the supplemental appendix. 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. 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, 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.

Figures 1 and 2 provide the time-varying volatility estimates obtained with the expectational updates. Specifically, the red lines in each figure provide the full-sample (smoothed) estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model’s volatility estimates, as well as real-time estimates of stochastic volatility (black lines). The real-time estimates are obtained by looping over time and estimating a historical volatility path at each forecast origin; these estimates underlie the forecast results considered in the next section. Note that, to improve chart readability, we reduce the number of panels on each page by omitting the estimates for the three-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. The cyclical pattern appears smaller for inflation, except that CPI inflation spiked sharply around the time of the Great Recession, presumably due to the dramatic, unexpected falloff in inflation that occurred as commodity prices collapsed.
- The overall magnitude of volatility for the nowcast horizon versus the expectational updates for longer horizons varies by variable, probably reflecting data timing. For growth and both measures of inflation, the level of volatility at the nowcast horizon exceeds the level of volatility at longer horizons. However, for the unemployment rate and T-bill rate, nowcast volatility is lower than longer-horizon update volatility, probably because the quarterly nowcast is often or always formed with the benefit of one month of data on the quarter.
- For the most part, for the period since the 1980s, the contours of SV estimates for inflation in the GDP price index and CPI are similar. There are of course some differences, including the relatively sharp late-2000s 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 real-time estimates. One dimension of this smoothness is that the real-time estimates tend to respond to recessions with a little delay; around recessions, the full sample estimates rise sooner than do the real-time estimates. In addition, in the case of CPI inflation, the late-2000s rise in volatility is larger in 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 3 and 4 report the forecast errors for each variable along with one-standard-deviation intervals, one set (in blue) obtained with the constant variance approach applied to a 60-observations rolling window of forecast errors and the other (in red) obtained from our stochastic volatility model of η_t . Again, for readability, we omit from the charts the estimates for the three-step-ahead horizon. Figures 3 and 4 provide a read on time variation in the width of confidence intervals and the accuracy of the two approaches.

The charts of the time paths of confidence intervals display the following broad patterns:

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

variance estimates) and then widen significantly (again, more so than the constant variance estimates) with the recessions of 2001 and 2007-2009. For most of the sample, the intervals are narrower with the SV approach than with the constant variance approach; however, this pattern does not so generally apply to CPI inflation.

- Across horizons, the contours of the confidence intervals (for a given approach) are very similar. With the SV model-based estimates, the similarities across horizons are particularly strong for horizons 1 through 4.¹⁴ 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. More broadly, with these variance estimates reflecting forecast uncertainty, as uncertainty varies over time, that uncertainty likely affects all forecast horizons, in a way captured by these SV estimates.

The coverage rates reported in Table 1 quantify the accuracy of the one-standard-deviation intervals shown in Figures 3 and 4. These show that the intervals based on our stochastic volatility model are consistently more accurate than the intervals based on the constant variance approach applied to forecast errors. Although we cannot claim that the SV-based approach yields correct coverage in all cases, it does so in the large majority of cases; the gap between the empirical and nominal rate is significant only in the case of TBILL forecasts at horizons $h = 0, 1,$ and 4 and RGDP forecasts at horizon $h = 4$. Moreover, the SV-based approach typically improves on the alternative approach, which in most cases yields coverage rates above 68 percent, reflecting bands that are too wide. For example, for GDP growth, the SV-based coverage rates range (across horizons) from 68.4 percent to 75.8 percent, whereas the constant-variance-based rates range from 77.6 percent to 79.6 percent, with all five departures from 68 percent large enough to be statistically significant.

¹⁴Note 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.

For the T-bill rate, the SV-based rates are much lower than the constant-variance-based rates at forecast horizons of 2 quarters or more — e.g., at the 2-quarter horizon, 70.2 percent with SV versus 84.5 percent for the constant variance baseline. For the inflation measures considered, results for the GDP price index are comparable to those for real GDP. But for CPI inflation, the coverage rates obtained with our SV model are broadly similar to those obtained with the constant variance benchmark approach.

To provide a broader assessment of density forecast accuracy, Table 2 reports the average CRPS. To simplify comparison, the table reports the level of the CRPS obtained with the constant variance approach and the percentage improvement in the CRPS of the SV-based forecasts relative to the constant-variance-based forecasts. For all variables, our SV model consistently offers density accuracy gains over the constant variance specification. The gains are largest for the T-bill rate, ranging from 7 to 14 percent. For GDP growth, the gains are still healthy, ranging from 3 to 9 percent. The gains in CRPS accuracy over the benchmark are statistically significant for growth and the T-bill rate. For the unemployment rate, the gains are smaller but significant at most horizons. For the inflation measures, the gains are still smaller and not statistically significant, but consistently positive, ranging from 1 to 3 percent. As noted above, although some studies have found modestly larger density gains associated with SV, these studies typically commingle benefits to point forecasts with benefits to the variance aspect of the density forecasts. In our case, the point forecasts are the same across the approaches, so any gains in density accuracy come entirely from variance-related aspects of the forecast distribution.

5.3 Out-of-sample results for VAR-SV specification

In the interest of brevity, in examining the efficacy of extending our baseline SV model to the VAR-SV specification, we present the out-of-sample results and omit figures with the full-sample VAR-SV estimates of volatility. The full-sample estimates for the VAR-SV model are qualitatively similar to the baseline SV estimates. Tables 3 and 4 provide one-standard-deviation coverage rates and CRPS values for the VAR-SV model, with comparison to the baseline constant forecast error

variance approach (repeating these results from Tables 1 and 2 for convenience).

The coverage rates reported in Table 3 show the intervals based on the VAR-SV model to be modestly more accurate than the intervals based on the constant variance approach applied to forecast errors. In broad terms, the advantages of the VAR-SV model over the benchmark constant variance case can be seen in the number of asterisks, with fewer statistically significant departures from correct coverage. As examples, the VAR-SV model yields coverage rates much lower than the constant variance benchmark for the unemployment and T-bill rates. However, in most cases, the advantages of the VAR-SV model are smaller than those of the baseline SV model. In most cases, coverage rates are higher with the VAR-SV model than with the baseline SV model. This is associated with less accurate coverage in most cases.

For broader density forecast accuracy, the CRPS averages provided in Table 4 show the VAR-SV specification to be useful for some variables and not others. For GDP growth, the unemployment rate, and the T-bill rate, the VAR-SV model yields density forecasts more accurate than those obtained with the benchmark constant variance approach, with gains up to 7 percent for growth, up to 12 percent for unemployment, and up to 30 percent for the T-bill rate. For the inflation variables, the VAR-SV model yields density forecasts modestly less accurate than the benchmark. When compared to the baseline SV model, the extension provided by the VAR-SV model is somewhat helpful for unemployment and T-bill forecasts (boosting the CRPS noticeably) and somewhat harmful for the other variables.

On balance, this evidence suggests that extending our baseline SV model to depart from its MDS assumption has a mixed payoff. It helps along some, but not all, dimensions. This finding suggests that the pre-whitening of multi-step forecast errors provided by the accounting identity used to obtain our baseline model is largely sufficient, although there are some variables for which adding VAR dynamics is a useful supplement to the baseline pre-whitening. The extensions of the VAR-SV specification seem to be most helpful for those series — the unemployment and T-bill rates — that exhibit (in results not presented for brevity) the largest degrees of bias or serial correlation in their expectational updates.

5.4 Additional robustness checks

In this subsection, we briefly summarize the robustness of our results with respect to five other changes in specification; the supplemental appendix provides additional details.

First, we have 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-observations setting of our baseline results. Lengthening to 80 observations the rolling window underlying the benchmark constant variance approach does not alter the picture we painted above: the constant variance approach commonly yields coverage rates in excess of the nominal rate of 68 percent. In addition, with the change in the rolling window length, it remains the case that our SV specification offers consistent gains to CRPS accuracy over the constant variance approach. Similarly, shortening the rolling window to 40 observations does not materially change the picture provided by the baseline results, although in forecast coverage, it slightly reduces the advantage of our SV-based model. Density accuracy as measured by the CRPS is only modestly affected by shortening the rolling window from 60 to 40 observations; our SV model-based approach maintains the same consistent advantage described above.

Second, we have also considered an alternative to the FE-CONST benchmark. This alternative also relies on a rolling window, but of the expectational updates and not the forecast errors directly. That is, it uses a more parametric approach assuming a time-invariant normal distribution for the nowcast error and the expectational updates collected in η_t (while maintaining the martingale difference sequence assumption): $\eta_t \sim N(\mathbf{0}, \Sigma)$. We employ Bayesian methods to estimate this model within the real-time setup described above, assuming a diffuse inverse-Wishart prior. Apart from nowcast uncertainty, the use of the expectational updates to estimate forecast error variances with rolling windows of data improves slightly on the FE-CONST benchmark, more so in the CRPS results than in the coverage results. Our proposed approach that incorporates stochastic volatility still offers consistent gains over this alternative rolling window benchmark (based on the expectational updates). In an overall sense, our methodological innovation has two components, one of which is the use of the expectational updates and the other is the use of stochastic volatility

(with the former enabling the latter), and both components appear helpful for the problem at hand.

Third, we have considered a version of our model with volatilities restricted to follow a single common factor process. In this case, the expectational updates at each forecast horizon of the data vector feature stochastic volatility, but the volatility process is common across horizons. As noted above, our baseline model estimates feature significant comovement of volatility; this restricted alternative imposes perfect comovement. In a forecasting sense, gains from parsimony might make such a restriction helpful to accuracy even if it is not entirely correct. However, results with the SPF forecasts are broadly similar to those from our baseline model. Compared to our baseline stochastic volatility model, imposing the single-factor structure does not yield consistently better or worse coverage rates or CRPS.

Fourth, we have considered an extension of the model to include multiple variables at once. More specifically, we considered a joint model for the three variables — with five horizons, this model totals 15 variables — for which we have data back to 1969: GDP growth, the unemployment rate, and GDP inflation. This specification permits correlation across the nowcast errors and expectational updates of different variables. On this dimension, too, our baseline findings appear to be robust. Broadly, coverage and CRPS results from this trivariate specification are similar to those from our baseline analysis. In coverage, the trivariate specification performs a little worse than our baseline, and in CRPS, it is a little better in some cases and worse in others.

Finally, we applied a version of a generalized VAR model with SV — described in Section 3.3 above — directly to data on observed forecast errors rather than expectational updates, henceforth referred to as FE-VAR(p)-SV. In light of the overlapping forecast windows, forecast errors should have stronger serial correlation than data on expectational updates and over longer lags, and we estimated this model variant using lag-length choices of $p = 2$ as well as $p = 5$. Compared to the constant variance benchmark, the FE-VAR(p)-SV model fares somewhat better in terms of coverage rates and forecast density accuracy (though not uniformly). But the FE-VAR(p)-SV model is generally inferior to our preferred SV or VAR-SV models that use expectational updates, η_t , as input data (with or without the MDS assumption). In most cases, coverage rates are higher (less

accurate) with the FE-VAR(p)-SV model than with the baseline SV model; by the CRPS measure, the FE-VAR(p)-SV model is less accurate than the baseline SV model for most, although not all, variables.

We leave as a subject for future research another form of a multivariate extension: making use of forecasts from multiple sources. In Reifschneider and Tulip (2007, 2017) and the Federal Reserve’s SEP, forecast accuracy is estimated by averaging the RMSEs 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 η_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.

6 Conclusions

Motivated in part by central bank fan charts that use historical forecast errors to quantify the uncertainty around forecasts, this paper develops a multiple-horizon specification of stochastic volatility for forecast errors from sources such as the SPF, the Blue Chip Consensus, or the Fed’s Greenbook, for the purpose of improving the accuracy of uncertainty estimates around the forecasts. Our approach can be used to form confidence bands around forecasts that allow for variation over time in the width of the confidence bands; the explicit modeling of the time variation of volatility eliminates the need for somewhat arbitrary judgments of sample stability.¹⁵

At each forecast origin, we have available the forecast error from the previous quarter and forecasts for the current quarter and the subsequent four quarters. To address the challenge of overlap in forecast errors across horizons, we formulate the model to make use of the current quarter (period t) nowcast error and the forecast updates for subsequent quarters (forecasts made in period t less forecasts made in period $t - 1$). These observations reflect the same information as the

¹⁵Examples of fan charts generated by our approach as well as the constant variance benchmark can be found in the supplemental appendix.

set of forecast errors for all horizons. However, unlike the vector of forecast errors covering multi-step horizons, the vector containing the forecast updates is serially uncorrelated, under conventional assumptions that the forecasts represent a vector of conditional expectations. For this vector of observations, we specify a multiple-horizon stochastic volatility model that can be estimated with Bayesian MCMC methods. From the estimates, we are able to compute the time-varying conditional variance of forecast errors at each horizon of interest.

Estimates of the model with the full sample of forecasts display considerable historical variation in forecast error variances, at each forecast horizon. Consistent with evidence from the VAR and DSGE literatures, the forecast error variances shrink significantly with the Great Moderation and tend to rise — temporarily — with each recession, most sharply for the recent Great Recession. To assess the performance of our approach in out-of-sample forecasting, we assess forecast coverage rates and the accuracy of density forecasts as measured by the continuous ranked probability score. We show that, by these measures, our proposed approach yields forecasts more accurate than those obtained using sample variances computed with rolling windows of forecast errors as in approaches such as those in Reifschneider and Tulip (2007, 2017). Admittedly, the choice between approaches involves some tradeoffs: our proposed approach offers a sophisticated and general way to identify and accommodate changes in forecast error variances, with modest benefits to accuracy, whereas the rolling window approach is somewhat simpler. Further work with other forecast sources, time periods, and model extensions would help to shed more light on such tradeoffs.

7 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.
- Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2015), “The Measurement and Characteristics of Professional Forecasters’ Uncertainty,” *Journal of Applied Econometrics* 30, 1029-1046.

- Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016), "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business and Economic Statistics* 34, 375-390.
- Chiu, Ching-Wai (Jeremy), Haroon Mumtaz, and Gabor Pinter (2017), "Forecasting with VAR Models: Fat Tails and Stochastic Volatility," *International Journal of Forecasting* 33, 1124-1143.
- Clark, Todd E. (2011), "Real-time Density Forecasts from BVARs with Stochastic Volatility," *Journal of Business and Economic Statistics* 29, 327-341.
- Clark, Todd E. and Francesco Ravazzolo (2015), "Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics* 30, 551-575.
- Clements, Michael P. (2014), "Forecast Uncertainty Ex Ante and Ex Post: U.S. Inflation and Output Growth," *Journal of Business and Economic Statistics* 32, 206-216.
- Clements, Michael P. (2016), "Are Macroeconomic Density Forecasts Informative?" Discussion Paper ICM-2016-02, Henley Business School, University of Reading.
- Cogley, Timothy, and Thomas J. Sargent (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post-World War II U.S.," *Review of Economic Dynamics* 8, 262-302.
- Coibion, Olivier, and Yuriy Gorodnichenko (2015), "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review* 105, 2644-2678.
- Croushore, Dean (2010), "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data," *BE Journal of Macroeconomics: Contributions* 10, Article 10.
- Croushore, Dean, and Tom Stark (2001), "A Real-Time Data Set for Macroeconomists," *Journal of Econometrics* 105, 111-130.
- Curdia, Vasco, Marco Del Negro, and Daniel Greenwald (2015), "Rare Shocks, Great Recessions," *Journal of Applied Econometrics* 29, 1031-1052.
- D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic Forecasting and Structural Change," *Journal of Applied Econometrics* 28, 82-101.

- D'Amico, Stefania, and Athanasios Orphanides (2008), "Uncertainty and Disagreement in Economic Forecasting," Finance and Economics Discussion Series Working Paper 2008-56, Federal Reserve Board of Governors.
- Del Negro, Marco, and Giorgio E. Primiceri (2015), "Time Varying Structural Vector Autoregressions and Monetary Policy: A Corrigendum," *Review of Economic Studies* 82, 1342-1345.
- Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13, 253-263.
- Diebold, Francis X., Frank Schorfheide, and Minchul Shin (2017), "Real-Time Forecast Evaluation of DSGE Models with Stochastic Volatility," *Journal of Econometrics* 201, 322-332.
- Durbin, J., and S.J. Koopman (2002), "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika* 89, 603-615.
- Giordani, Paolo, and Paul Soderlind (2003), "Inflation Forecast Uncertainty," *European Economic Review* 47, 1037-1059.
- Giordani, Paolo, and Mattias Villani (2010), "Forecasting Macroeconomic Time Series with Locally Adaptive Signal Extraction," *International Journal of Forecasting* 26, 312-325.
- Gneiting, Tilmann, and Adrian E. Raftery (2007), "Strictly Proper Scoring Rules, Prediction, and Estimation," *Journal of the American Statistical Association* 102, 359-378.
- Gneiting, Tilmann, and Roopesh Ranjan (2011), "Comparing Density Forecasts Using Threshold and Quantile Weighted Proper Scoring Rules," *Journal of Business and Economic Statistics* 29, 411-422.
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (2004), "Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors," *Journal of Econometrics* 122, 185-212.
- Justiniano, Alejandro and Giorgio E. Primiceri (2008), "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review* 93, 604-641.
- Kenny, Geoff, Thomas Kostka, and Federico Masera (2014), "How Informative Are the Subjective Density Forecasts of Macroeconomists?" *Journal of Forecasting* 33, 163-185.
- Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic Volatility: Likelihood

- Inference and Comparison with ARCH Models,” *Review of Economic Studies* 65, 361-393.
- Knoppel, Malte (2014), “Efficient Estimation of Forecast Uncertainty Based on Recent Forecast Errors,” *International Journal of Forecasting* 30, 257-267.
- Knoppel, Malte, and Guido Schultefrankfeld (2012), “How Informative Are Central Bank Assessments of Macroeconomic Risks?” *International Journal of Central Banking* 8, 87-139.
- Krueger, Fabian, Sebastian Lerch, Thordis L. Thorarinsdottir, and Tilmann Gneiting (2017), “Probabilistic Forecasting and Comparative Model Assessment Based on Markov Chain Monte Carlo Output,” manuscript, November.
- Lahiri, Kajal, and Xuguang Sheng (2010), “Measuring Forecast Uncertainty by Disagreement: The Missing Link,” *Journal of Applied Econometrics* 25, 514-538.
- Mertens, Elmar, and James M. Nason (2018), “Time-varying Stickiness in Professional Inflation Forecasts,” BIS Working Paper 713, Bank for International Settlements.
- Newey, Whitney K., and Kenneth D. West (1987), “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica* 55, 703-708.
- _____ (1994), “Automatic Lag Selection in Covariance Matrix Estimation,” *The Review of Economic Studies* 61, 631-653.
- Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima, (2007), “Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference,” *Journal of Econometrics* 140, 425-449.
- Primiceri, Giorgio E. (2005), “Time Varying Structural Vector Autoregressions and Monetary Policy,” *Review of Economic Studies* 72, 821-852.
- 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.
- _____ (2017), “Gauging the Uncertainty of the Economic Outlook Using Historical Forecasting Errors: The Federal Reserve’s Approach,” FEDS Working Paper 2017-20, Federal Reserve Board of Governors.

Rossi, Barbara, and Tatevik Sekhposyan (2014), “Evaluating Predictive Densities of US Output Growth and Inflation in a Large Macroeconomic Data Set,” *International Journal of Forecasting* 30, 662-682.

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

——— (2016), “Core Inflation and Trend Inflation,” *Review of Economics and Statistics* 98, 770-784.

Tulip, Peter and Stephanie Wallace (2012), “Estimates of Uncertainty Around the RBA’s Forecasts,” Research Discussion Paper 2012-07, Reserve Bank of Australia.

Table 1: Forecast error coverage rates: one-standard-deviation bands

Variable	Forecast horizon					eval. begin
	0	1	2	3	4	
Panel A: SV						
RGDP	72.06	69.63	73.13	68.42	75.76*	1983:Q4
UNRATE	70.80	70.59	65.93	61.19	62.41	1983:Q4
PGDP	73.53	71.11	71.64	70.68	71.97	1983:Q4
CPI	72.09	70.59	65.48	68.67	68.29	1996:Q3
TBILL	76.74*	77.65*	70.24	63.86	50.00**	1996:Q3
Panel B: FE-CONST						
RGDP	77.94***	78.52**	77.61*	78.95*	79.55**	1983:Q4
UNRATE	72.99	82.35***	85.19***	87.31***	86.47***	1983:Q4
PGDP	75.00*	77.04**	77.61**	78.20**	79.55***	1983:Q4
CPI	72.09	64.71	69.05	67.47	71.95	1996:Q3
TBILL	79.07*	88.24***	84.52**	80.72	79.27	1996:Q3

Note: The table reports the empirical out-of-sample coverage rates of one-standard-deviation bands. The sample uses predictions made from the date given in the right-most column through 2017:Q4 (and realized forecast errors as far as available). The upper panel provides results based on our proposed multi-horizon SV model. The lower panel provides results based on the FE-CONST model estimated over rolling windows with 60 quarterly observations. Statistically significant departures from a nominal coverage of 68% (as predicted under a normal distribution) are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 2: Density forecast accuracy as measured by CRPS

Variable	Forecast horizon					eval. begin
	0	1	2	3	4	
RGDP						
(SV rel.)	3.01%**	7.50%***	7.96%***	9.27%***	7.58%***	1983:Q4
(FE-CONST)	0.82	1.02	1.10	1.16	1.17	
UNRATE						
(SV rel.)	1.78%*	2.82%**	3.56%**	3.44%*	2.25%	1983:Q4
(FE-CONST)	0.08	0.17	0.25	0.34	0.43	
PGDP						
(SV rel.)	1.03%	1.41%	1.83%	2.59%	3.00%	1983:Q4
(FE-CONST)	0.50	0.56	0.60	0.63	0.68	
CPI						
(SV rel.)	1.98%	2.35%	1.49%	1.63%	2.35%	1996:Q3
(FE-CONST)	0.66	1.05	1.09	1.10	1.10	
TBILL						
(SV rel.)	11.36%***	13.99%***	13.00%***	9.85%**	6.86%	1996:Q3
(FE-CONST)	0.07	0.23	0.40	0.58	0.76	

Note: The table reports CRPS results for out-of-sample density forecasts. The sample uses predictions made from the date given in the right-most column through 2017:Q4 (and realized forecast errors 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 rather than FE-CONST; positive numbers indicate improvement of SV over the FE-CONST case. The bottom row reports the CRPS for the FE-CONST case, which has been estimated over rolling windows with 60 quarterly observations. Statistical significance of the differences in average CRPS — assessed with a Diebold and Mariano (1995) test — is indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Table 3: Forecast error coverage rates: one-standard-deviation bands, VAR-SV specification

Variable	Forecast horizon					eval. begin
	0	1	2	3	4	
Panel A: VAR-SV						
RGDP	74.26	73.33	76.12*	72.93	77.27*	1983:Q4
UNRATE	66.42	76.47*	75.56	74.63	72.18	1983:Q4
PGDP	73.53	74.07	79.85***	77.44**	79.55***	1983:Q4
CPI	67.44	72.94	67.86	71.08	78.05***	1996:Q3
TBILL	69.77	81.18**	75.00	65.06	65.85	1996:Q3
Panel B: FE-CONST						
RGDP	77.94***	78.52**	77.61*	78.95*	79.55**	1983:Q4
UNRATE	72.99	82.35***	85.19***	87.31***	86.47***	1983:Q4
PGDP	75.00*	77.04**	77.61**	78.20**	79.55***	1983:Q4
CPI	72.09	64.71	69.05	67.47	71.95	1996:Q3
TBILL	79.07*	88.24***	84.52**	80.72	79.27	1996:Q3

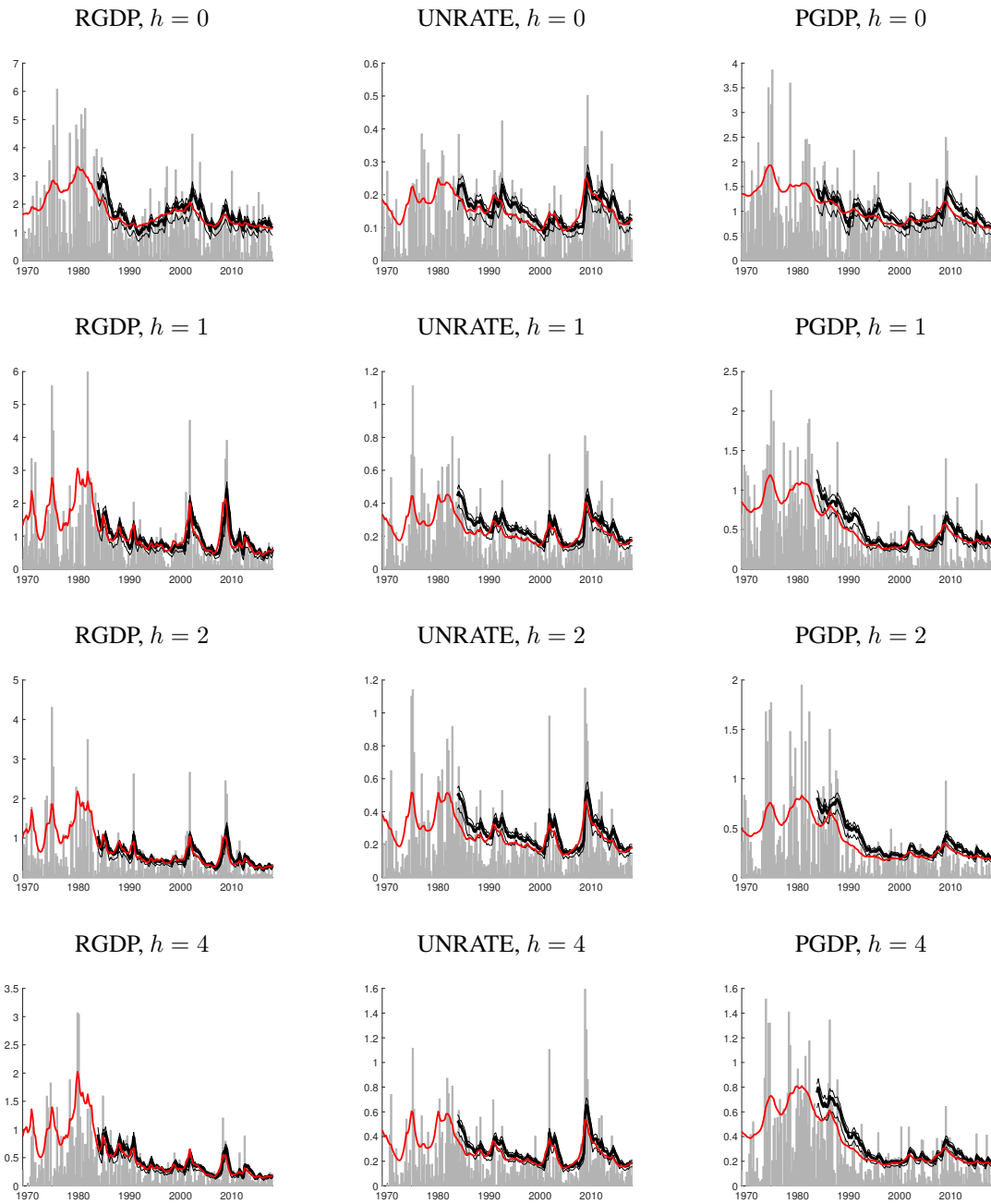
Note: The table reports the empirical out-of-sample coverage rates of one-standard-deviation bands. The sample uses predictions made from the date given in the right-most column through 2017:Q4 (and realized forecast errors as far as available). The upper panel provides results based on our proposed multi-horizon VAR-SV model. The lower panel provides results based on the FE-CONST model estimated over rolling windows with 60 quarterly observations. Statistically significant departures from a nominal coverage of 68% (as predicted under a normal distribution) are indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

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

Variable	Forecast horizon					eval. begin
	0	1	2	3	4	
RGDP						
(VAR-SV rel.)	1.77%	6.77%***	7.36%***	7.07%***	4.80%**	1983:Q4
(FE-CONST)	0.82	1.02	1.10	1.16	1.17	
UNRATE						
(VAR-SV rel.)	12.15%***	11.34%***	10.36%***	9.02%**	6.22%	1983:Q4
(FE-CONST)	0.08	0.17	0.25	0.34	0.43	
PGDP						
(VAR-SV rel.)	-2.29%	-2.86%	-3.00%	-3.45%	-5.92%	1983:Q4
(FE-CONST)	0.50	0.56	0.60	0.63	0.68	
CPI						
(VAR-SV rel.)	9.86%**	-0.41%	-1.79%	-2.84%	-3.53%	1996:Q3
(FE-CONST)	0.66	1.05	1.09	1.10	1.10	
TBILL						
(VAR-SV rel.)	29.74%***	25.45%***	25.40%***	23.76%***	21.21%***	1996:Q3
(FE-CONST)	0.07	0.23	0.40	0.58	0.76	

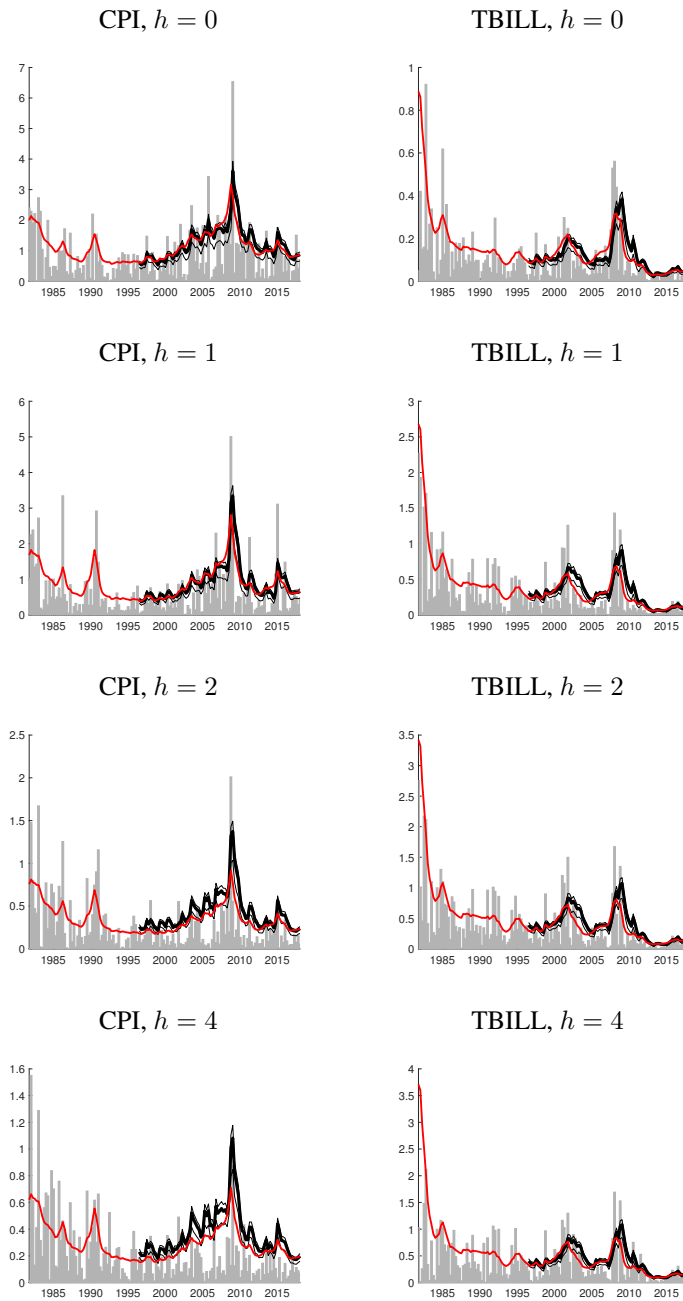
Note: The table reports CRPS results for out-of-sample density forecasts. The sample uses predictions made from the date given in the right-most column through 2017:Q4 (and realized forecast errors as far as available). For each variable, the top row reports the relative CRPS calculated as the percentage decrease of the CRPS when using VAR-SV rather than FE-CONST; positive numbers indicate improvement of VAR-SV over the FE-CONST case. The bottom row reports the CRPS for the FE-CONST case, which has been estimated over rolling windows with 60 quarterly observations. Statistical significance of the differences in average CRPS — assessed with a Diebold and Mariano (1995) test — is indicated by *, **, or ***, corresponding to 10, 5, and 1 percent significance, respectively.

Figure 1: Stochastic Volatility in Expectational Updates for Real GDP Growth, the Unemployment Rate, and GDP Price Index Inflation



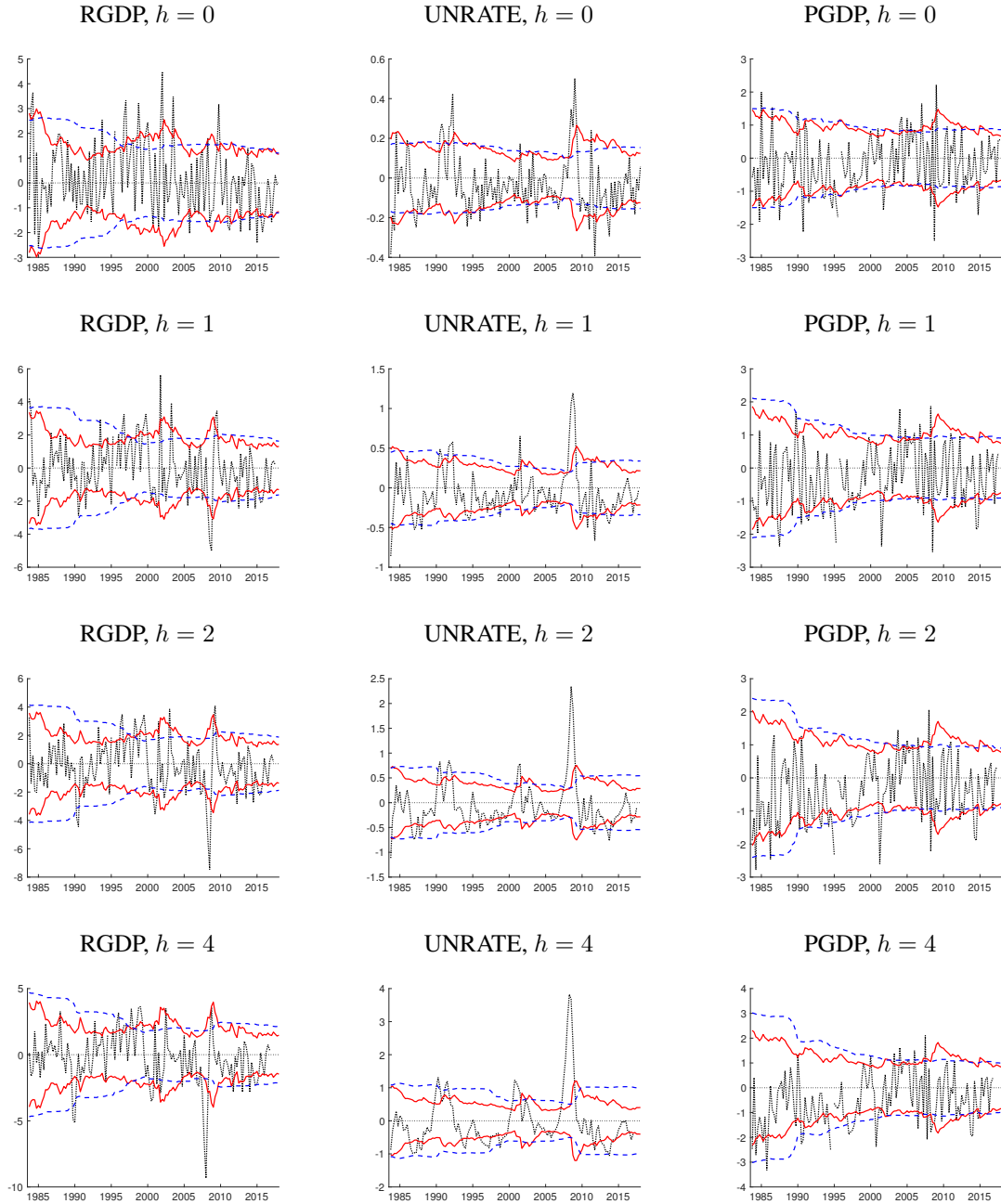
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond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

Figure 2: Stochastic Volatility in Expectational Updates for CPI Inflation and the T-bill Rate



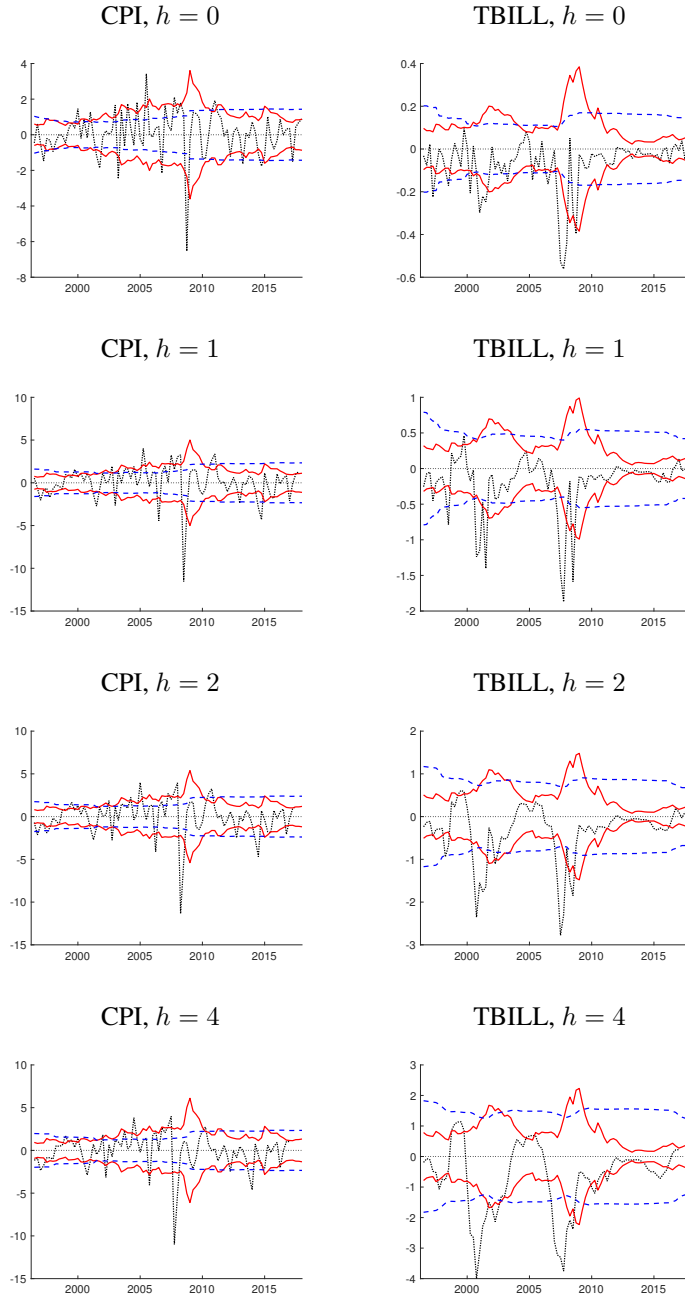
Note: The red lines in each figure provide the full-sample estimates of stochastic volatility (reported as standard deviations, or $\lambda_{i,t}^{0.5}$ in the model notation). For comparison, the figures include (in gray bars) the absolute values of the expectational updates, which roughly correspond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

Figure 3: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for Real GDP Growth, the Unemployment Rate, and GDP Price Index Inflation



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

Figure 4: Ex-Ante Uncertainty Bands and Ex-Post Forecast Errors for CPI Inflation and the T-bill Rate



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