Federal Reserve Bank of Cleveland Working Paper Series

Constructing Fan Charts from the Ragged Edge of SPF Forecasts

Todd E. Clark, Gergely Ganics, and Elmar Mertens

Working Paper No. 22-36R

August 2024

Suggested citation: Clark, Todd E., Gergely Ganics, and Elmar Mertens. 2024. "Constructing Fan Charts from the Ragged Edge of SPF Forecasts." Working Paper No. 22-36R. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202236r.

Federal Reserve Bank of Cleveland Working Paper Series
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.

See more working papers at: www.clevelandfed.org/research. Subscribe to email alerts to be notified when a new working paper is posted at: https://www.clevelandfed.org/subscriptions.

Constructing Fan Charts from the Ragged Edge of SPF Forecasts*

Todd E. Clark,[1] Gergely Ganics,[2] and Elmar Mertens[3]

[1] *Federal Reserve Bank of Cleveland,* [2] *Banco de España,* [3] *Deutsche Bundesbank*

July 24, 2024

Abstract

We develop models that take point forecasts from the Survey of Professional Forecasters (SPF) as inputs and produce estimates of survey-consistent term structures of expectations and uncertainty at arbitrary forecast horizons. Our models combine fixed-horizon and fixed-event forecasts, accommodating time-varying horizons and availability of survey data, as well as potential inefficiencies in survey forecasts. The estimated term structures of SPF-consistent expectations are comparable in quality to the published, widely used short-horizon forecasts. Our estimates of time-varying forecast uncertainty reflect historical variations in realized errors of SPF point forecasts, and generate fan charts with reliable coverage rates.

Keywords: Term structure of expectations, uncertainty, survey forecasts, fan charts

JEL classification codes: E37, C53

---

*Parts of this paper were circulated earlier under the title "Constructing the Term Structure of Uncertainty from the Ragged Edge of SPF Forecasts" and were completed while G. Ganics was with the Central Bank of Hungary. We gratefully acknowledge helpful suggestions and comments received from Cem Çakmakı, Refet Gürkaynak, Ed Knotek, Mike McCracken, James Mitchell, Tom Stark, and workshop or conference participants at the NBER Summer Institute 2022, 2022 IAAE conference, 2022 Dolomiti Macro Meeting, 2022 NBER-NSF SBIES conference, 2022 NBER-NSF Time Series conference, 2022 Conference on Real-Time Data Analysis, Methods, and Applications, Heidelberg University, and Deutsche Bundesbank. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, the Federal Reserve System, the Banco de España, the Deutsche Bundesbank, or the Eurosystem. Replication files are available at `https://github.com/elmarmertens/ClarkGanicsMertensSPFfancharts`.

1 Introduction

Both economic policymaking and forecasting research commonly use the macroeconomic projections of professional forecasters. Such forecasts with long histories include the (US) Survey of Professional Forecasters (SPF), Blue Chip Consensus, and Consensus Economics, as well as Federal Reserve forecasts published in the Tealbook or Greenbook and the Federal Open Market Committee's (FOMC) Summary of Economic Projections (SEP). Typically, the availability of such forecasts is uneven across forecast horizons. For example, the SPF includes both (1) fixed-horizon quarterly point forecasts, at shorter horizons, and (2) fixed-event annual forecasts (i.e., forecasts of a given event, in this case an economic outcome in a specific calendar year, made at different points in time before the event (Nordhaus, 1987)), available at longer horizons. Forecasters, policy analysts, and researchers regularly consider forecast uncertainty; many central banks (e.g., the Federal Reserve and the European Central Bank) publish estimates of forecast uncertainty.

In this paper, we develop models that take published survey point forecasts as inputs and produce estimates of a longer, more complete term structure (across horizons) of survey-consistent point forecasts, along with a term structure of forecast uncertainty. Our applications use the SPF, and our methods can exactly replicate its quarterly forecasts at short horizons while interpolating through the longer-horizon forecasts from the SPF's annual predictions. Our estimates of forecast uncertainty reflect dispersion in past errors from the SPF's point forecasts, with time-varying uncertainty captured through stochastic volatility (SV).[1]

Matching the term structure of SPF forecasts with predictions from a simple time series model of the outcome variable typically requires measurement error, but that is not the case in our models.[2] Specifically, we assume that, beyond a given horizon, forecasts are equal to a drifting mean, while forecasts up to that horizon follow a flexible moving-average structure that can match arbitrary term structures of expectations. Nevertheless, to avoid excessively volatile imputation that can arise from small inconsistencies between quarterly and annual SPF predictions, we model annual forecasts with fat-tailed measurement errors that are a priori close to zero.

Our approach extends Clark, McCracken, and Mertens (2020) (hereafter, CMM) in a few directions. First, we model an SPF-consistent term structure of quarterly expectations that extends arbitrarily far beyond the four quarters covered in CMM. Second, our inference on expectations beyond the near term is informed by the SPF's fixed-event (annual) forecasts that extend up to three years ahead. Doing so requires handling the so-called ragged edge of the SPF, whose quarterly publication leads to systematic variations in the fixed-event horizons of annual forecasts. Third, we develop a generalized model that allows for bias (both conditional and unconditional) in the SPF. Time-varying bias in SPF forecasts and other departures from full rationality have been highlighted in studies including Bianchi, Ludvigson, and Ma (2022), Coibion and Gorodnichenko (2015), and Farmer, Nakamura, and Steinsson (2024). We provide a rich and flexible model of the joint dynamics between outcomes and expectational data provided by the SPF.

Practically speaking, what might one do with a more complete term structure of expectations and uncertainty? As one application, our approach can be used to produce SPF-consistent fan charts patterned after those from the FOMC's SEP, which report calendar-year forecasts of Q4/Q4 growth rates of GDP and prices and the Q4 level of the unemployment rate. The width of the uncertainty bands in these SEP fan charts is two times the historical root mean square errors (RMSEs) obtained using a few different forecasts computed over 20-year rolling windows, as developed in Reifschneider and Tulip (2019). To illustrate, from a forecast origin of 2024Q1 Figure 1 provides SEP-style fan charts with SPF-consistent calendar-year forecasts of GDP growth, CPI inflation, and the unemployment rate. The fan charts are generated from our estimates of predictive densities that are conditioned on the SPF. For the figure, we then transform draws from these predictive densities into calendar-year predictions for 2024-2027 that match the data conventions of the SEP. The fan charts include a point forecast and a 68 percent forecast uncertainty band. For comparison, the charts also include the corresponding projections (red dotted lines) from the March 2024 SEP. The SPF-consistent point forecasts show GDP growth remaining near 2 percent for the next few years, with unemployment near 4 percent. The SPF forecasts have CPI inflation slowing from 2024 to 2025 and changing little in 2026 and 2027. The SEP's point forecasts are similar, although the SEP includes a further reduction in inflation in 2026. But, in terms of forecast uncertainty, the

SPF-consistent and SEP fan charts differ more materially: While the 68 percent forecast interval for the SPF is reasonably wide, the SEP's is notably wider (in keeping with a broader historical pattern that we document later in the paper).

[Figure 1 about here.]

In the body of this paper, we flesh out the approach that underlies this example. We first show that our baseline model yields SPF-consistent quarterly real-time forecasts of GDP growth, unemployment, and inflation that match the published fixed-horizon quarterly forecasts from the SPF and interpolate through the annual fixed-event point forecasts. The model's estimates of forecast uncertainty vary over time, temporarily rising around recessions, and generally rise with the forecast horizon. Our second set of results addresses the quality of the extended SPF forecasts obtained from our models (both the baseline specification that treats SPF forecasts as rational and the generalized model that allows departures from full rationality), applying a variety of metrics to out-of-sample forecasts. Our estimates indicate that the generalized model captures reasonably well the empirical extent of non-rationality emphasized by Coibion and Gorodnichenko (2015) and subsequent studies. Nonetheless, the SPF forecasts obtained from our restricted baseline model and the generalized version perform comparably in various metrics of forecast quality. More specifically, our estimated forecasts for the SPF at forecast origin t are efficient predictions of the forecasts published in the SPF at origin $t + 1$. Out to forecast horizons as long as 16 quarters ahead, quarterly forecasts from the baseline model and its generalization are similar in point and density accuracy. The two models also perform comparably in unconditional coverage rates, yielding forecast confidence bands that are reasonably accurate, but not perfect. Tests of the uniformity of probability integral transforms indicate that our estimated predictive densities of SPF forecasts are correctly calibrated. Together, we take these results as evidence that the quality of our SPF-consistent estimated forecasts is comparable to the quality of the published short-horizon forecasts, which are widely seen as useful in research and practice.

The paper proceeds as follows. Section 2 provides an overview of the related literature on survey forecasts not covered above or in subsequent sections. Section 3 describes the SPF forecasts and data used. Section 4 presents our models and briefly describes estimation. Section 5 provides results. Section 6 concludes. A supplementary online appendix provides additional results.

2 Related Literature

Over the course of our analysis, we will consider the rationality, calibration, and accuracy of the term structure of SPF-consistent forecasts. A long literature has examined the same properties of published fixed-event and fixed-horizon survey forecasts. More specifically, some studies have examined whether professional forecasts display properties consistent with optimal (typically, under quadratic loss) forecasts and rational expectations. In one example, Patton and Timmermann (2012) develop new rationality tests based on rationality restrictions taking the form of bounds on second moments of the data across forecast horizons and apply them to forecasts from the Federal Reserve's Greenbook. As regards accuracy, focusing on inflation forecasts, Ang, Bekaert, and Wei (2007) find survey forecasts hard to beat by a battery of forecasting methods.

While we focus on professional forecasts in and of themselves, another long literature has sought to use professional forecasts to improve forecasts from time series models. Faust and Wright (2009) use professional forecasts as jumping-off points for models to improve the accuracy of forecasts from time series models. Wright (2013) shows that the forecast accuracy of a Bayesian vector autoregression (VAR) can be improved by using long-run survey forecasts as priors on the model's steady states. Banbura, et al. (2021), Ganics and Odendahl (2021), and Krüger, Clark, and Ravazzolo (2017) improve forecasts from Bayesian VARs through entropic tilting toward survey-based nowcasts or forecasts. Frey and Mokinski (2016) instead add survey-based nowcasts as endogenous variables in Bayesian VARs, using priors so that the dynamics of the survey forecasts inform the parameter estimates of the dynamics of the actual data. In a similar vein, Doh and Smith (2022) develop priors that align a VAR's (a priori) forecasts with survey predictions.

While our approach uses SPF forecasts as the only inputs to obtain a term structure of forecasts and uncertainty, a number of other studies have estimated term structures of survey forecasts by using a time series model for y_t that generates forecasts $E_t y_{t+h}$ and assuming that observed survey forecasts are equal to model-implied forecasts plus a measurement error, $F_t y_{t+h} = E_t y_{t+h} + \text{noise}_{t+h}$. This approach is employed by studies such as Aruoba (2020), Crump, et al. (2023), Grishchenko, Mouabbi, and Renne (2019), Kozicki and Tinsley (2012), and Mertens and Nason (2020). Crucially, in applications such as these, the use of measurement error is necessary, as it serves as fitting error in matching term structures of a tightly parameterized model to the SPF data. In our application, we attach measurement error to annual SPF forecasts out of concern about (small) inconsistencies in the SPF data; while our model can perfectly replicate the SPF without measurement error, its presence avoids excessively volatile imputations.

Potential drawbacks of this approach that features a time series model of the data include the attribution of part of the observed term structure of survey expectations to measurement error and the imposition of a (typically low-order) time series model on the term structure of "true" expectations ($E_t y_{t+h}$). For some applications, such an approach might provide a potentially beneficial form of shrinkage. In addition, the approach can be used to pool surveys from different sources to extract a common set of underlying expectations. However, the measurement error approach comes at the cost of discarding part of the survey respondents' judgment and broader modeling. Instead, as explained below, one of our models aims to fit model-implied forecasts to survey forecasts, $F_t y_{t+h} = E_t y_{t+h}$, and our more general model sees differences between forecasts from the SPF and the model as bias (and thus predictable forecast errors, instead of measurement error).

Complementing our work on a rich term structure of forecast uncertainty several years into the future, many other studies have used the fixed-event forecasts of professionals to examine forecast uncertainty and its variation over time. For example, with data on fixed-event forecasts from Consensus Economics, Patton and Timmermann (2011) use an unobserved components model to examine the predictability of growth and inflation across different forecast horizons and measure average forecast uncertainty by mean square forecast errors.

3 Data

Because the availability of forecasts in the SPF informs aspects of our model, we detail the data in this section before taking up the model in Section 4. We examine quarterly and annual forecasts from the SPF for a basic set of major macroeconomic aggregates commonly included in research on the forecasting performance of models such as VARs or dynamic stochastic general equilibrium models: real GDP growth (RGDP), the unemployment rate (UNRATE), and inflation in the GDP price index (PGDP) and consumer price index (CPI).[3] The SPF forecasts are widely studied, publicly available, and the longest time series of forecasts for a range of variables.

We obtained the SPF forecasts from the Federal Reserve Bank of Philadelphia's Real-Time Data Research Center. In all cases, we form the point forecasts using the average over all SPF responses. Reflecting the data available, our estimation samples start with 1969Q1 for GDP growth, unemployment, and GDP inflation and 1981Q4 for CPI inflation; the last forecast origin for which we evaluate out-of-sample forecasts is 2023Q4, and we also discuss fan charts conditioned on the 2024Q1 SPF. At each forecast origin, the available fixed-horizon point forecasts typically span five quarters, from the current quarter through the next four quarters. Since 1981Q3, the SPF has included fixed-event point forecasts for the current and next calendar year. In 2005Q3, the forecast horizon for CPI inflation was extended to include annual forecasts for one additional year, i.e., covering up to two calendar years ahead, and in 2009Q2, the forecast horizon was similarly extended for GDP growth and unemployment to extend up to three years ahead.

Given the fixed-event nature of the SPF's calendar-year forecasts, the number of quarters until the end of the longest-horizon forecast varies over the course of a year. For example, the 2021Q4 SPF included fixed-event annual forecasts of growth and unemployment for the current year and the next three, so that the last annual forecast extends 12 quarters ahead (the annual forecast reported for 2024 includes 2024Q4, 12 quarters beyond the 2021Q4 forecast origin). In the 2022Q1 SPF, the last annual forecast for 2025 includes 2025Q4, 15 quarters beyond the forecast origin. These variations in the SPF's effective forecast horizon are also known as the "ragged edge." Our methods allow us to consistently construct fixed-horizon term structures that, at every quarterly forecast origin, extend arbitrarily far beyond the ragged edge.

The SPF uses certain conventions in its forecasts that we incorporate in the measurement specification of our model, as described further in Section 4. In terms of the SPF data we use, predictions for the unemployment rate are expressed directly as quarterly and annual-average levels (depending on the forecast horizon), which can be mapped directly into our model. Similarly, for CPI inflation, the forecasts are reported as percentage changes in quarterly and Q4/Q4 index levels that are used as such in our model. However, for GDP and its price index, the SPF data files provide forecasts in levels, which we convert into growth rates. Specifically, depending on the forecast horizon, point forecasts pertain to quarterly or annual-average levels, which we convert to growth rates based on information included in the survey. For quarterly forecast targets, we use the lagged quarterly level as the basis. To obtain the next-year forecast of annual-average growth, we use the SPF's predictions for the current year as base values (and analogously for the forecasts two and three years ahead). For each forecast horizon, we calculate growth rates for GDP and the GDP price index as differences in log levels, as our model uses log-linearized expressions for quarterly and annual-average growth in these variables that are detailed in Section 4.

To estimate our model, we also need measures of the outcomes of the variables. In the case of GDP growth and GDP inflation, for which data can be substantially revised over time, 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 Philadelphia Fed's Real-Time Data Set for Macroeconomists (RTDSM). For forecast evaluation, we measure the outcomes of GDP growth and GDP inflation with the RTDSM vintage published two quarters after the outcome date.[4] Because revisions to quarterly data are relatively small for the unemployment rate and CPI inflation, we simply use the historical time series available in the St. Louis Fed's FRED database to measure the outcomes and corresponding forecast errors for these variables.

4 Model and Estimation

In broad terms, our models can be seen as multivariate unobserved components models. We design the state space specifications to match arbitrary term structures of expectations, with application to the SPF in this paper. In all cases, we specify and estimate the models on a variable-by-variable basis (i.e., separately for GDP growth, unemployment, and each inflation measure).

This section begins by spelling out the general setup of our models, including our underlying assumptions. We then proceed with detailing the transition equations of our state space models, starting with a specification that treats forecast updates as martingale differences and then proceeding to a more general specification that allows for biases and persistence in forecast errors and expectations updates. After laying out transition dynamics, we present the measurement equation included in both models, followed by our specification of the innovation components of the transition equations. Throughout, we use $\boldsymbol{Y_t}$ to denote the state vector (partially latent) of quarterly forecasts and $\boldsymbol{Z_t}$ to denote the measurement vector of the available SPF point forecasts.

4.1 General setup

Throughout, y_t refers to either (1) the annualized quarterly log growth rate of GDP or its price index, (2) a quarterly average level of the unemployment rate, or (3) the annualized quarterly (simple) percent change in the CPI. We denote SPF-consistent forecasts of the scalar outcome y_{t+h} made in period t by $F_t y_{t+h}$ and rational (or unbiased) forecasts as $E_t y_{t+h}$. "SPF-consistent" means that these forecasts are either directly observed from the SPF or values imputed from our model under the assumption that those values represent the (average) SPF response. The vector $\boldsymbol{Y_t}$ — the "SPF-consistent" term structure of expectations — includes the lagged realized value, y_{t-1}, and forecasts from horizons $h = 0, \ldots, H$:

$$\boldsymbol{Y_t} \equiv \left[y_{t-1},\ F_t y_t,\ F_t y_{t+1}, \ldots, F_t y_{t+h}, \ldots, F_t y_{t+H} \right]'. \tag{1}$$

Note that, with $y_{t-1} = F_t y_{t-1}$, we have $F_t \boldsymbol{Y_t} = \boldsymbol{Y_t}$. To derive the transition equation that captures the dynamics of $\boldsymbol{Y_t}$, we begin with definitional equations for the (lagged) nowcast error, e_{t-1}, as well as SPF forecast updates, $(F_t - F_{t-1})y_{t+h}$:[5]

$$y_{t-1} = F_{t-1}y_{t-1} + e_{t-1}\,, \tag{2}$$

$$F_t y_{t+h} = F_{t-1}y_{t+h} + (F_t - F_{t-1})y_{t+h}, \quad \forall\, h \geq 0\,. \tag{3}$$

In each of these equations, the term on the left is an element of $\boldsymbol{Y_t}$, whereas the first term on the right is an element of $\boldsymbol{Y_{t-1}}$, which suggests a strategy to derive a recursion for the dynamics of $\boldsymbol{Y_t}$.

We also make the following three assumptions.

1. The term structure of SPF-consistent forecasts is flat beyond horizon H, with y_t^* denoting the endpoint of the term structure of SPF forecasts:

$$y_t^* \equiv F_t y_{t+H+1} = F_t y_{t+H+j} \quad \forall\, j > 0\,. \tag{4}$$

2. The endpoint y_t^* is an unbiased (rational) long-run forecast of y_t, corresponding to the Beveridge-Nelson trend in y_t and following a random walk process:

$$y_t^* \equiv \lim_{j\to\infty} E_t y_{t+H+j} = y_{t-1}^* + w_t^* \;\Rightarrow \quad E_{t-1}w_t^* = 0. \tag{5}$$

3. The endpoint of the SPF term structure, y_t^*, is the common trend in all elements of $\boldsymbol{Y_t}$ and deviations from trend, also called "gaps" and denoted by $\boldsymbol{\tilde{Y}_t}$, are (mean) stationary:

$$\boldsymbol{Y_t} = \boldsymbol{\tilde{Y}_t} + \boldsymbol{1}y_t^*\,, \quad \text{with} \quad \lim_{h\to\infty} E_t\boldsymbol{\tilde{Y}_{t+h}} = \boldsymbol{\bar{Y}}\,, \tag{6}$$

where the mean vector for the gaps, $\boldsymbol{\bar{Y}}$, captures unconditional bias in the SPF at different horizons (detailed further in the supplementary online appendix), and $\boldsymbol{1}$ denotes a vector of ones. With this assumption, we are treating SPF forecasts as being at least weakly rational, so that their forecast errors are stationary, which implies that SPF forecasts at different horizons are cointegrated among each other and with the outcome variable (Mertens, 2016).[6]

Together with equation (3), these assumptions imply the following recursive law of motion for the vector of gaps $\tilde{\boldsymbol{Y}}_{\boldsymbol{t}}$ (where $\boldsymbol{I}$ is the identity matrix):

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{t}} = \left(\boldsymbol{I} - \tilde{\boldsymbol{\Psi}}\right)\bar{\boldsymbol{Y}} + \tilde{\boldsymbol{\Psi}}\tilde{\boldsymbol{Y}}_{\boldsymbol{t-1}} + \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}\,, \tag{7}$$

$$\text{with} \quad \tilde{\boldsymbol{\Psi}} = \begin{bmatrix} 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 1 \\ 0 & \dots & \dots & \dots & \dots & 0 \end{bmatrix}, \qquad \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} \equiv \begin{bmatrix} e_{t-1} \\ (F_t - F_{t-1})y_t \\ (F_t - F_{t-1})y_{t+1} \\ \vdots \\ (F_t - F_{t-1})y_{t+h} \\ \vdots \\ (F_t - F_{t-1})y_{t+H} \end{bmatrix} - \boldsymbol{1}w_t^* - \bar{\boldsymbol{\eta}}\,, \tag{8}$$

and $\bar{\boldsymbol{\eta}} \equiv \left(\boldsymbol{I} - \tilde{\boldsymbol{\Psi}}\right)\bar{\boldsymbol{Y}}$. All eigenvalues of $\tilde{\boldsymbol{\Psi}}$ are zeros, so that $\tilde{\boldsymbol{\Psi}}$ is a stable matrix.[7] To close the description of the state equation, we need to model the dynamics of $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ with particular attention to its persistence, i.e., $E_{t-1}\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$, as well as its distribution, which we model via variations in $\text{Var}_{t-1}\,\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$. Below we explain the two different state equation specifications considered and the shock process.

The first assumption above also has implications for the setting of H when implementing our model. While the SPF provides forecasts up to three calendar years ahead, its longer-run forecasts are stated in terms of annual predictions that straddle several quarters. Thus, to match the observed SPF, our state vector does not have to extend all the way to the end of the third calendar year ($H = 15$). Instead, we can utilize our assumption that the term structure of SPF-consistent forecasts is flat beyond some horizon H and describe a three-year-ahead forecast with a state space extending only to, say, $H = 12$. For concreteness, consider the case of a forecast, collected in Q1, that should match the average of the four quarters of the third year ahead. With $H = 12$, only the last element of $\boldsymbol{Y}_{\boldsymbol{t}}$ points into the third year ahead, and forecasts for quarters two to four of the third year are set identical to the trend, with zero gaps.

In this spirit, we generally choose H such that when the forecast origin is in Q1, H points to the first quarter of the farthest annual horizon covered by the SPF forecasts. An exception is made for data covering only SPF forecasts up to the next year, where H is set to 5 (instead of 4), since the observed fixed-horizon SPF forecasts already extend through $h = 4$. As part of our out-of-sample forecast analysis, models are (re-)estimated over different sub-samples of data, and we adjust H accordingly. Specifically, we set $H = 5$ for data samples on GDP growth and unemployment prior to 2009Q2, as well as for CPI inflation prior to 2005Q4, and inflation in the GDP price index (entire sample). For estimates covering data that include SPF forecasts up to three years ahead (GDP growth and unemployment since 2009Q2) we set $H = 12$, and for CPI inflation data since 2005Q4, which include SPF forecasts up to two years ahead, we set $H = 8$. Note that, even though the state vector ends with $F_t y_{t+H}$ and H is no larger than 12, our endpoint assumption allows us to simulate forecast densities arbitrarily far ahead, and we report densities up to 16 quarters ahead for all variables throughout.

4.2 Transition equation in the martingale case

Our first model assumes SPF forecasts are rational so that $F_t y_{t+h} = E_t y_{t+h} \; \forall \, h$. It follows that forecasts are unbiased, $\bar{\boldsymbol{Y}} = \boldsymbol{0}$, and that forecast updates are martingale difference sequences (MDS): $E_{t-1}\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} = 0$ and $E_{t-1} w_t^* = 0$. With these restrictions, the mean dynamics of the gap vector $\tilde{\boldsymbol{Y}}_{\boldsymbol{t}}$ are fully specified by equation (7), with the restriction $\bar{\boldsymbol{Y}} = \boldsymbol{0}$ imposed.

To describe the density of $\tilde{\boldsymbol{Y}}_{\boldsymbol{t}}$, we use a block-SV structure with fat-tailed shocks to $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ and w_t^* that will be described further below. Although this MDS-based model is written with a state vector $\boldsymbol{Y}_{\boldsymbol{t}}$ containing SPF forecasts, we should emphasize that, in this specification, we are not actually attributing a specific time series model to the evolution of SPF forecasts; we are taking the observed fixed-horizon and fixed-event forecasts as given and using a time series process to interpolate missing fixed-horizon forecasts out to H steps ahead. With the MDS specification, all we need is the historical evolution of the SPF up to the forecast origin t and previous nowcast errors in order to estimate SPF-consistent point forecasts and uncertainty bands around the forecasts; we do not need a time series model of y_t.

4.3 Transition equation of the VAR model

The MDS model places tight restrictions on the evolution of forecast updates by assuming that SPF forecasts are rational. Our second, more general specification drops that assumption. This model allows for (a) unconditional bias in the form of a non-zero intercept vector $\bar{\boldsymbol{Y}}$ in the transition dynamics of equation (7) and (b) conditional bias by modeling the updates in (detrended) forecasts, $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$, as a VAR(1):[8]

$$\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} = \tilde{\boldsymbol{\Pi}}\,\tilde{\boldsymbol{\eta}}_{\boldsymbol{t-1}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}\,, \qquad \text{with} \quad \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \sim \mathcal{N}(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}}), \tag{9}$$

and details for $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}}$ to be described further below. In keeping with the common-trend assumption, the transition matrix $\tilde{\boldsymbol{\Pi}}$ is assumed to be stable. It follows that the state equation for the vector of gaps $\tilde{\boldsymbol{Y}}_{\boldsymbol{t}}$ takes a VAR(2) form:

$$\tilde{\boldsymbol{Y}}_{\boldsymbol{t}} = \left(\boldsymbol{I} - \tilde{\boldsymbol{\Psi}}\right)\left(\boldsymbol{I} - \tilde{\boldsymbol{\Pi}}\right)\bar{\boldsymbol{Y}} + \left(\tilde{\boldsymbol{\Psi}} + \tilde{\boldsymbol{\Pi}}\right)\tilde{\boldsymbol{Y}}_{\boldsymbol{t-1}} - \left(\tilde{\boldsymbol{\Psi}}\,\tilde{\boldsymbol{\Pi}}\right)\tilde{\boldsymbol{Y}}_{\boldsymbol{t-2}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}\,. \tag{10}$$

The MDS model is nested in this specification, under the restrictions $\bar{\boldsymbol{Y}} = \boldsymbol{0}$ and $\tilde{\boldsymbol{\Pi}} = \boldsymbol{0}$. Absent these restrictions, the model allows for unconditional and conditional bias in the SPF. With that, the VAR model still captures in $\boldsymbol{Y}_{\boldsymbol{t}}$ an SPF-consistent term structure of expectations that matches the observed SPF data (exactly for quarterly SPF forecasts, and up to measurement error for the annual forecasts). But predictions for the outcome variables generated by the VAR model will generally differ from $\boldsymbol{Y}_{\boldsymbol{t}}$, whereas point forecasts generated by the MDS specification will not.

4.4 Measurement equation

To explain the measurement equation of our state space models, we define:

$$\bar{y}_t = 1/4 \cdot \sum_{j=0}^{3} y_{t-j}, \tag{11}$$

and
$$\hat{y}_t \equiv 100 \times \log\left(\frac{I_t + I_{t-1} + I_{t-2} + I_{t-3}}{I_{t-4} + I_{t-5} + I_{t-6} + I_{t-7}}\right), \tag{12}$$

$$\approx 1/16 \cdot \left(y_t + 2 \cdot y_{t-1} + 3 \cdot y_{t-2} + 4 \cdot y_{t-3} + 3 \cdot y_{t-4} + 2 \cdot y_{t-5} + y_{t-6}\right), \tag{13}$$

where I_t denotes the quarterly level of GDP or its price index.

With these definitions of multi-period forecast targets, we can cover the different types of annual forecasts published in the SPF: When t corresponds to a date in Q4, $\bar{y}_t$ represents the annual average level of the unemployment rate as well as Q4/Q4 percent changes in the CPI, and $\hat{y}_t$ captures percent changes in annual average levels of GDP and its price index. As discussed in the supplementary online appendix, the formula for $\hat{y}_t$ in (13) represents a log-linear approximation to the growth rate in average levels of calendar years ending at t and $t-4$ in (12).[9] The measurement equations of our models, detailed below, use the approximation in (13) to relate annual SPF forecasts for GDP growth and GDP price inflation to forecasts (or lagged realizations) of y_t.[10]

We denote survey forecasts collected at forecast origin t for targets y_{t+h}, $\bar{y}_{t+h}$, and $\hat{y}_{t+h}$, by $F_t y_{t+h}$, $F_t \bar{y}_{t+h}$, and $F_t \hat{y}_{t+h}$, respectively. At time t, the SPF provides observations of $F_t y_{t+h}$, $F_t \bar{y}_{t+h}$, and/or $F_t \hat{y}_{t+h}$, for different (but separate) values of $h \geq 0$. The measurement vector $\boldsymbol{Z_t}$ contains the available observations from the SPF as well as a reading of the last realized value, y_{t-1}, that is available to SPF respondents at time t. In addition to the SPF's quarterly fixed-horizons forecasts, $F_t y_{t+h}$ for $h = 0, 1, \ldots, 4$, we include in $\boldsymbol{Z_t}$ the available readings of fixed-event forecasts for the next year and beyond.[11] For example, with t = 2024Q1, the measurement vector for GDP growth is $\boldsymbol{Z_t} = \left[y_{t-1},\ F_t y_t,\ F_t y_{t+1}, \ldots, F_t y_{t+4}, F_t \hat{y}_{t+7}, F_t \hat{y}_{t+11}, F_t \hat{y}_{t+15}\right]'$. For t = 2024Q2, the annual forecast entries become $F_t \hat{y}_{t+6}$, $F_t \hat{y}_{t+10}$, and $F_t \hat{y}_{t+14}$. The corresponding measurement vectors for the unemployment rate differ only by using annual forecasts expressed in terms of $\bar{y}_t$ instead of $\hat{y}_t$. Further details are described in the supplementary online appendix.

Our framework can match a term structure of expectations to an arbitrary set of observable SPF data. It could do so without a need for measurement noise. However, while not essential, attaching measurement error to at least some SPF observations may be warranted and helpful. It is possible that the reported readings of annual forecasts contain some small discrepancies, due, for example, to the computation of growth rates using GDP forecasts reported in levels with rounding. Consider, for example, the case of SPFs in the third quarter of each year that include an annual forecast for the next-year SPF. Given the handful of quarterly forecasts published in the SPF, to the model the only "free" quarterly forecast for matching the annual forecast is the projection for the fourth quarter of next year. In the applications to GDP growth and inflation in the GDP deflator, that quarterly forecast has a weight of 1/16 in the annual approximation in (13), which in turn means that potential discrepancies get scaled up by a factor of 16 in the imputation of the fourth quarter forecast for next year. As detailed in the supplementary online appendix, these mechanics can (and do) lead to excessively volatile imputations. Accordingly, in our implementation, the measurement equations of our MDS and VAR models add measurement error to the annual forecasts, with separate measurement error processes for each quarter of the year. Formally, we partition the measurement vector into two parts, and assume that only the first part is observed without error:

$$\boldsymbol{Z_t} = \begin{bmatrix} \boldsymbol{Z_{q,t}} \\ \boldsymbol{Z_{a,t}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C_{q,t}} \\ \boldsymbol{C_{a,t}} \end{bmatrix} \boldsymbol{Y_t} + \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{n_t} \end{bmatrix}, \quad n_{i,t} \sim \mathcal{N}(0, \sigma^2_{i,t}), \tag{14}$$

where $n_{i,t}$ is the ith element of $\boldsymbol{n_t}$ and denotes the measurement error for the ith element of $\boldsymbol{Z_{a,t}}$. The loading matrices $\boldsymbol{C_{q,t}}$ and $\boldsymbol{C_{a,t}}$ vary over time as a function of the available measurements, and contain fixed coefficients of 0, 1, or fractional values as indicated in the measurement definitions (11) and (13). The measurement errors are independent across elements of $\boldsymbol{Z_{a,t}}$, modeled as conditionally Gaussian with zero mean and time-varying variance, as detailed below.

4.5 Shock distributions

This section describes our modeling of the shocks in the MDS and VAR specifications, starting with the distribution of $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}$, the shocks to the gaps vector of our state space model. (In the MDS case, $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} = \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}$.) Throughout, we model these shocks as conditionally Gaussian: $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \sim \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}})$.

We model variations in uncertainty through a time-varying $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}}$. To do so, we build on the specification of Chan (2020) in which a scalar factor affects the entire Gaussian shock vector, imparting perfectly correlated variations in uncertainty to all shocks. We extend his framework by splitting the shock vector into multiple blocks, each with its own common volatility factor. In our case, the shock vector reflects shocks to forecasts at different horizons, and the block structure allows us to model distinct variations in uncertainty at different forecast horizons. Specifically, we adopt a structure with two blocks, such that $\tilde{\boldsymbol{\varepsilon}}_{\mathbf{1},t}$ consists of shocks to the lagged realized outcome and forecasts up to $h = 3$ and $\tilde{\boldsymbol{\varepsilon}}_{\mathbf{2},t}$ consists of shocks to forecasts for $h = 4, \ldots, H$. The cutoff at $h = 3$ between blocks reflects the desire for the first block to mostly capture forecasts for quarters within the current year, while the second block captures forecasts for quarters in the next year and beyond.[12] Specifically, we adopt the following block-SV structure:

$$\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} = \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_{\mathbf{1},t} \\ \tilde{\boldsymbol{\varepsilon}}_{\mathbf{2},t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \tilde{\boldsymbol{K}} \\ 0 & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varepsilon}^*_{\mathbf{1},t} \\ \boldsymbol{\varepsilon}^*_{\mathbf{2},t} \end{bmatrix}, \quad \text{with} \begin{bmatrix} \boldsymbol{\varepsilon}^*_{\mathbf{1},t} \\ \boldsymbol{\varepsilon}^*_{\mathbf{2},t} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{1,t} \cdot \tilde{\boldsymbol{\Sigma}}_{\mathbf{11}} & 0 \\ 0 & \lambda_{2,t} \cdot \tilde{\boldsymbol{\Sigma}}_{\mathbf{22}} \end{bmatrix} \right), \tag{15}$$

in which $\tilde{\boldsymbol{K}}$ is a matrix (with dimension $4 \times (H-2)$) of coefficients to be estimated. This SV structure yields the following time-varying variance-covariance matrix of the cyclical shocks:

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{t}} = \begin{bmatrix} \boldsymbol{I} & \tilde{\boldsymbol{K}} \\ 0 & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \lambda_{1,t} \cdot \tilde{\boldsymbol{\Sigma}}_{\mathbf{11}} & 0 \\ 0 & \lambda_{2,t} \cdot \tilde{\boldsymbol{\Sigma}}_{\mathbf{22}} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \tilde{\boldsymbol{K}} \\ 0 & \boldsymbol{I} \end{bmatrix}'. \tag{16}$$

This block-SV representation includes a deliberate upper-triangular structure to capture covariance between the shocks to the first and second blocks of shocks. Under this structure, the volatilities of longer-horizon forecasts — latent states, not directly observed forecasts, for $h \geq 4$ — are driven by a single common, longer-horizon factor and not impacted by volatilities at shorter horizons. Volatilities of shorter-horizon forecasts — which are directly observed — are allowed to be impacted by both the common, shorter-horizon factor and the longer-horizon factor. We think of time-varying volatility as having medium-frequency business cycle drivers as well as some higher frequency drivers; the former impacts forecast volatility at all horizons, whereas the latter only impacts the volatilities of shorter-horizon forecasts.

It is commonly understood that the ordering of variables affects estimates of VARs with SV processes for each variable (see, e.g., Arias, Rubio-Ramirez, and Shin (2023)). With our two-block specification, within each block, the common SV specification means that variable ordering does not matter. However, ordering has some impact through the block-triangular structure; changing to a lower-triangular rather than upper-triangular structure would change the model estimates. In this setting, though, for the reasons given above for the upper-triangular structure, this specification choice is less arbitrary than the simple ordering of variables in a conventional VAR with SV.

The scalar factors $\lambda_{1,t}$ and $\lambda_{2,t}$ impart time variation and fat tails to the shock vector $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}$. Building on, among others, Carriero, et al. (2022), Chan (2020), and Jacquier, Polson, and Rossi (2004), we model these factors as the products of iid inverse-gamma draws and persistent stochastic volatility processes:

$$\lambda_{i,t} = \phi_{i,t} \cdot \tilde{\lambda}_{i,t} \,, \qquad \forall\, i = 1, 2 \,, \tag{17}$$

$$\text{with} \quad \phi_{i,t} \sim \mathcal{IG}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right) , \qquad \log \tilde{\boldsymbol{\lambda}}_{\boldsymbol{t}} \equiv \begin{bmatrix} \log \tilde{\lambda}_{1,t} \\ \log \tilde{\lambda}_{2,t} \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{bmatrix} \log \tilde{\boldsymbol{\lambda}}_{\boldsymbol{t-1}} + \boldsymbol{\epsilon}_{\boldsymbol{t}}^{\boldsymbol{\lambda}} \,, \tag{18}$$

and $\boldsymbol{\epsilon}_{\boldsymbol{t}}^{\boldsymbol{\lambda}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Phi})$. The iid inverse-gamma draws add fat tails in the form of a multivariate t distribution with ν_i degrees of freedom to each block. The vector SV process $\log \tilde{\boldsymbol{\lambda}}_{\boldsymbol{t}}$ has correlated shocks and is normalized to a mean of zero, obviating the need for normalizing assumptions on the constant-coefficient matrices $\tilde{\boldsymbol{\Sigma}}_{\mathbf{11}}$ and $\tilde{\boldsymbol{\Sigma}}_{\mathbf{22}}$.

To complete the models, we need to specify processes for the trend shock (w_t^*) and measurement error ($n_{i,t}$) for each annual forecast i. We use (independent) horseshoe specifications for these shocks. Carvalho, Polson, and Scott (2010) proposed the horseshoe for modeling sparse regressions, i.e., regressions with many regressors, many of which are expected to be irrelevant, with only a few attracting substantial mass a posteriori. So while a horseshoe prior places considerable mass on coefficient values of zero, it also has particularly fat tails to generate (few) sizable coefficient estimates. Similarly, our application of the horseshoe to shocks reflects our presumption that most realizations of shocks are close to zero while some can instead be sizable. To allow the size of measurement errors to vary systematically within the year, we apply separate horseshoe processes to forecasts collected in different quarters of the year. The horseshoe has a conditionally Gaussian representation, with conditional mean 0 and a time-varying conditional variance, as indicated in equation (14) for the case of the noise shocks. Similarly, for the trend shocks, we have:

$$w_t^* \sim \mathcal{N}(0, \omega_t^2). \tag{19}$$

In modeling ω_t^2 and $\sigma_{i,t}^2$, the horseshoe representation employs global and local shrinkage (i.e., specific to each variable and each t, respectively) as detailed in the supplementary online appendix.

4.6 Estimation

We estimate the models using Bayesian Markov chain Monte Carlo (MCMC) methods — specifically, a Gibbs sampler. The model estimation is conditioned on joint data for observed realizations and SPF predictions for a given economic variable (like GDP growth), but estimated separately for different economic variables.

In the MDS specification, the objects to be estimated include the ρ_i and $\boldsymbol{\Phi}$ parameters of the SV processes, the constant innovation covariance matrix $\tilde{\boldsymbol{\Sigma}}$, the parameters of the horseshoe models for innovations to the long-run forecast and measurement errors, the time series of the latent volatility states in $\tilde{\boldsymbol{\lambda}}_t$, and the time series of latent forecast states contained in $\boldsymbol{Y}_t$. We sample the parameters of the SV processes using a conventional Gaussian prior and conditional posterior for ρ_i and an inverse Wishart prior and conditional posterior for $\mathbf{\Phi}$. We sample $\tilde{\mathbf{\Sigma}}$ with an inverse Wishart prior and conditional posterior. We estimate the volatility states in $\tilde{\boldsymbol{\lambda}}_t$ with the mixture approach of Kim, Shephard, and Chib (1998). Sampling of the horseshoe components in shocks to trend and noise follows the MCMC scheme described in Makalic and Schmidt (2016). Estimation of the VAR requires additional steps in the Gibbs sampler to draw the coefficients $\tilde{\mathbf{\Pi}}$ from their multivariate normal conditional posterior, while accounting for the heteroskedasticity in the shock processes. Further details are provided in the supplementary online appendix.

Sampling the latent forecast states contained in $\boldsymbol{Y}_t$ involves more complexity. Since our model assumes no measurement error in at least some of its observables, we face an ill-defined posterior precision that cannot be directly handled by conventional precision-based samplers. To efficiently draw the latent forecast states in $\boldsymbol{Y}_t$, we build on a new precision-based sampler developed in Mertens (2023) with details described in the supplementary online appendix. We retain 3,000 draws after a burn-in sample of 3,000 initial draws. When simulating the model's predictive density we sample 100 paths of future realizations of stochastic volatility and other state variables for each MCMC draw, resulting in $S = 300,000$ predictive density draws.

5 Results

This section begins with results on the term structures of expectations and forecast uncertainty. We focus on SPF-consistent forecasts from our MDS model specification. We then examine various aspects of the quality of out-of-sample forecasts, from both our MDS and VAR specifications. In all of these results, we examine forecasts starting in 1990Q1. We conclude the section with some comparisons of our SPF-based forecasts against those of the FOMC as reported in the SEP.

5.1 Term structures of expectations and forecast uncertainty

To illustrate the term structures of SPF-consistent expectations produced by our MDS model, Figure 2 presents fan charts of quarterly forecasts from an origin of 2024Q1. As detailed above, to generate these out-of-sample forecasts the model uses as inputs quarterly SPF forecasts up through the four-step-ahead horizon and the available annual fixed-event forecasts (except for the current year), available as of the indicated forecast origin (and not future forecasts or other data). We report the mean forecasts and 68 and 90 percent forecast intervals.

[Figure 2 about here.]

In this example, the SPF-consistent point forecasts show GDP growth remaining around 2 percent, the unemployment rate edging up just a little before settling around 4 percent, and inflation eventually settling at or (in the case of the CPI) a little above 2 percent. As expected, the forecast intervals widen — i.e., forecast uncertainty increases — as the quarterly horizon grows. Below we take up in more detail the term structure of uncertainty and its variation over time.

Regarding the construction of the SPF-consistent quarterly forecasts, by design the MDS model's forecasts for horizons 0 through 4 quarters exactly match the reported SPF projections. Forecasts at horizons 5 through H interpolate (up to measurement error) through the SPF's published annual forecasts of growth, unemployment, and inflation. That interpolation is not necessarily simply linear; the model is capable of capturing richer dynamics based on the historical comovement among the observed updates in SPF forecasts. At these horizons of 5 and more quarters, there is some uncertainty around the model's estimates of the SPF-consistent quarterly forecast (i.e., uncertainty around the estimate of the point forecast — the state $F_t y_{t+h}$ that is latent for $h \geq 5$). Comparisons omitted in the interest of brevity confirm that, as expected, the uncertainty around the quarterly forecast estimates is reduced by having annual forecasts as measurements. However, the uncertainty around the latent state estimate is a small component of the overall forecast uncertainty reflected in the fan charts. The overall forecast uncertainty is captured and driven by our model's estimates of the time-varying variances of historical forecast errors and updates.

To provide a broader time perspective on the term structure of SPF-consistent expectations, Figure 3 shows the history of term structures of SPF-consistent forecasts (these are all out-of-sample) from 1990 through 2019. For readability, we only report forecasts originating in the first quarter of each year, and the figures stop in 2019 to avoid the pandemic period's volatility. A given colored solid line provides the SPF-consistent quarterly forecasts for a given forecast origin, with different lines for different origins, and dotted lines provide the outcomes of each variable. Across quarters for a given origin, the forecasts often show some fluctuations before settling at the model's take on the long-run mean. Over time, however, these means show some changes; the forecast paths evolve to follow the broad contours of economic outcomes. This pattern is starkest with the decline of actual unemployment and inflation from about 1990 to 2000. Also as expected, from one forecast origin to another, short-horizon forecasts change more abruptly than do the longer-horizon forecasts. The forecasts of growth and unemployment reflect the widely familiar difficulty of forecasting recessions in advance; for example, the SPF-consistent forecasts of growth turn sharply negative after the recessions of 1990-91 and 2007-2009 become evident and, in each case, anticipate recoveries. On the other hand, SPF forecasters were consistently surprised by how quickly unemployment declined after the 2007-2009 recession.

[Figure 3 about here.]

The dynamic behavior of the SPF-consistent forecasts across quarters reflects forecasters' views of underlying data processes. In results detailed in the supplementary online appendix, it is possible to derive univariate processes for the time series y_t implied by estimates of our model of SPF forecasts (conditioning on the entire SPF term structure). Essentially, using the model's steady-state Kalman filter and the Kalman gain, we can obtain moving average (MA) coefficients of the process for y_t. In MDS model-based estimates for GDP growth and the two inflation measures, the coefficient profiles resemble low-order MA processes, with a little more persistence for inflation than growth. Following a surprise increase, these variables steadily return to their longer-run levels. The implied process for the unemployment rate resembles a much more persistent AR process with hump-shaped dynamics, such that following a surprise increase, unemployment rises further before slowly declining. Estimated processes implied by the VAR specification are qualitatively similar to those from the MDS specification.

[Figure 4 about here.]

Turning from SPF-consistent point forecasts to the term structure of forecast uncertainty, Figure 4 depicts the term structure of uncertainty around quarterly forecasts, from 1990 to 2023. For constructing the figure, we measure uncertainty by the width of the 68 percent bands of the model's predictive densities estimated in real time (i.e., we report out-of-sample forecast uncertainty). For readability, the chart includes a subset of quarterly horizons, including a few short ones and longer horizons at increments of 4 quarters.

In these results, uncertainty is noticeably higher at longer horizons (8 or more quarters) than shorter horizons (0 to 4 quarters). While this applies to all variables, short-horizon uncertainty compared to long-horizon uncertainty is relatively high for GDP growth. The uncertainty of out-of-sample forecasts fluctuates significantly over time (a feature we capture by including stochastic volatility in the model). For GDP growth and unemployment, uncertainty rose some following the 2001 recession and increased more notably around the Great Recession and again a few years into the ensuing recovery (while omitted from the chart, our estimates also show uncertainty declining significantly with the Great Moderation that occurred earlier in the sample). Then the outbreak of COVID-19 produced an unprecedented, but temporary, spike in uncertainty in 2020. As of 2023Q4, forecast uncertainty for GDP growth and unemployment remains above its pre-pandemic level. The uncertainty estimates for the inflation measures sharply trended down from 1990 to 2000 (inflation uncertainty also declined significantly with the Great Moderation) and then fluctuated over the remainder of the sample, with some increases around the Great Recession and pandemic.

5.2 Quality of MDS and VAR forecasts

This section examines the quality of our SPF-consistent forecasts along various dimensions. This quality assessment includes out-of-sample forecasts from both the MDS and VAR models. Unless otherwise noted, our evaluations use forecasts through 2023Q4. (The supplementary online appendix contains corresponding results for a sample of forecasts ending before the pandemic, which yield the same qualitative findings.) After presenting the results from the MDS and VAR specifications, we discuss interpretations and implications of their relative performance.

Regarding departures from forecast rationality, before taking up forecast quality we assess the extent to which the general VAR specification reflects them. Building on the framework of Nordhaus (1987) and subsequent empirical work, as well as work on information rigidities in macroeconomics, Coibion and Gorodnichenko (2015) propose quantifying departures from rationality with the coefficients of regressions of forecast errors on forecast updates:

$$y_{t+h} - F_t y_{t+h} = \alpha_h + \beta_h \left(F_t - F_{t-1}\right) y_{t+h} + \text{error}_{t+h}. \tag{20}$$

Many subsequent studies — often pooling estimates across forecast horizons — have provided related evidence using this regression for various forecasts. We apply this metric to estimates of our VAR specification. From standard regression calculus, the β_h coefficient of (20) is a function of the moments of forecast errors and updates. As detailed in the supplementary online appendix, our VAR specification of SPF forecasts can be used to compute corresponding moment values using the estimated coefficients of the VAR and in turn to compute the implied β_h coefficient. We use this approach to obtain pooled estimates of β for the horizons for which observed fixed-horizon forecasts are available from the SPF.[13] As reported in the supplementary online appendix, these estimates are generally in line with the literature that finds positive coefficients of small-to-modest magnitudes, indicating some departures from full rationality in professional forecasts.

One metric of the quality of our estimated term structure of SPF (point) forecasts is whether they are good predictions of the SPF published in the subsequent quarter — in particular, whether our estimates of forecasts not directly observed at forecast origin t are good predictions of the forecasts (at the same horizon) directly observed in the SPF published at forecast origin $t+1$. We rely on estimates of regressions styled after the efficiency test of Mincer and Zarnowitz (1969):

$$F_{t+1} y_{t+h} = \alpha_h + \beta_h F_t y_{t+h} + \text{error}_{t+h}, \tag{21}$$

where we only use horizons such that the $t+1$ forecast on the left side is directly observed and the t forecast on the right relies — at longer horizons — on latent quarterly forecasts estimated from our model. (While we simplify notation here to just refer to quarterly forecasts, our regressions also cover annual forecasts at longer horizons.) If the predictions are good, the slope coefficient β_h will be close to 1, with an intercept α_h of 0. Table 1 reports these regression estimates for a sample of 1990-2023, separately by model, variable, and horizon.

Generally, these tests of the predictability of SPF point forecasts indicate that our models fare reasonably well in producing forecasts that are efficient or nearly efficient forecasts of the next and directly observed SPF. For example, for GDP growth, the slope coefficients obtained for the MDS forecasts are all close to 1; only at a horizon of $h = 4$ does the estimate differ significantly from unity. With MDS forecasts, the same pattern applies for UNRATE, PGDP, and CPI: most slope coefficients are close to 1, with one or two horizons departing significantly from a coefficient of unity. By this metric, the quality of the forecasts at longer horizons appears to be comparable to that at shorter horizons. The results are broadly similar for SPF forecasts obtained with the VAR generalization of the model, albeit with a few more rejections of slope coefficients of unity (most sharply for PGDP). Of course, given the count of tests reflected in the table's results for different variables and horizons, some rejections of unity slope coefficients are to be expected given Type II error. Broadly, we take these results as evidence that our interpolation of quarterly forecasts at longer horizons is yielding forecasts comparable in quality to the shorter-horizon forecasts.

[Table 1 about here.]

As another measure of forecast quality, we conduct an out-of-sample evaluation of real-time forecasts generated by the MDS and VAR versions of the model. Table 2 compares the accuracy of point forecasts, evaluated in terms of RMSE, as well as density predictions, evaluated by the continuous ranked probability score (CRPS). From 1990Q1 onward, out-of-sample forecasts are generated for quarterly horizons from $h = 0$ through 16, by re-estimating the model at each forecast origin (using all available data since 1968Q4), and simulating its predictive density. The results are reported as ratios with MDS results in the denominator, so that entries below (above) 1 indicate that the VAR forecasts are more (less) accurate than the MDS forecasts.

[Table 2 about here.]

These results indicate that the point and density forecasts (again, these are out-of-sample) from the MDS and VAR specifications have broadly similar accuracy over the full sample. For RGDP, the RMSE and CRPS ratios are all very close to 1, at both shorter horizons (for which SPF forecasts are directly observed) and longer horizons (for which quarterly forecasts must be interpolated by the models). For the inflation measures, the RMSE and CRPS ratios are typically just a bit above 1.00, giving a very slight advantage (not enough to be considered material) to the MDS model.[14] The MDS model's advantage over the VAR is modestly greater for UNRATE.

As another measure of forecast quality, Table 3 reports unconditional coverage rates — the frequency with which actual quarterly outcomes fall within 68 and 90 percent forecast intervals. (In the 90 percent case, there are relatively few observations available for evaluating accuracy in the tails of the distributions.) A frequency of more (less) than 68/90 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. To save space, the table reports results for just the VAR specification; the MDS model yields similar coverage rates, reported in the supplementary online appendix.

[Table 3 about here.]

Overall, these results indicate that our models applied to the SPF yield forecasts and densities that are reasonable, but not perfect, from the perspective of unconditional coverage. The coverage rates are best for CPI and PGDP inflation, with empirical coverage rates comparable to nominal rates and relatively few rejections of correct coverage. Coverage performance is more mixed for GDP growth. At longer horizons, coverage rates are close to nominal rates (with few significant departures), although empirical rates run a bit below nominal rates, indicating the intervals are a little too narrow. At shorter horizons, the empirical coverage rates are more notably below nominal rates. Unconditional coverage performance is also mixed for unemployment forecasts — for 68 percent coverage and less so for 90 percent coverage. For unemployment, 68 percent predictive intervals tend to be too wide at shorter horizons and too narrow at longer horizons.

Of course, unconditional coverage rates may be seen as a limited window into the calibration of the predictive densities. Rossi and Sekhposyan (2019) develop a broader approach to assessing whether predictive densities are correctly specified: testing the uniformity of the empirical cumulative distribution functions (CDFs) of probability integral transforms (PITs) of forecasts. Their approach compares empirical CDFs to the 45-degree line (representing a uniform CDF) with appropriate confidence intervals for departures from uniformity. In the interest of brevity, Figure 5 provides these results for MDS and VAR forecasts of GDP growth and unemployment, for selected horizons (the supplementary online appendix provides results for inflation).

The PITs comparisons indicate that the predictive densities are correctly calibrated for GDP growth, with empirical CDFs of the PITs within the confidence intervals. This applies to both the MDS and VAR forecasts. Importantly, the forecasts are correctly specified at both the shorter horizons for which quarterly SPF forecasts are directly observed and the longer horizons for which quarterly forecasts must be estimated with our models. The departures from uniformity are somewhat greater for the PITs of unemployment forecasts, with the empirical CDFs further away from the 45-degree line. Graphically, while these departures come close to or hit the confidence interval bounds, they generally do not fall outside the bounds, except in the case of $h = 4$ forecasts from the MDS model. In this instance, the VAR performs a little better than the MDS model by allowing some bias and rationality departures in the forecasts.

[Figure 5 about here.]

While we omit details in the interest of brevity (estimates are available in the supplementary online appendix), there is one metric by which the MDS specification has a clear advantage: Marginal data densities (the conventional Bayesian measure of model fit) estimated with one-step-ahead predictive likelihoods indicate that the MDS specification fits the data significantly better than the more general VAR model. Nonetheless, we have shown that the VAR model implies departures from full rationality in keeping with those documented in Coibion and Gorodnichenko

(2015) and subsequent studies, in which forecast updates of professional forecasters show statistically significant serial correlation. Yet the battery of additional metrics of forecast quality reported in this section indicate that the more restricted MDS specification is on par with the more general VAR; the models are hard to distinguish in those comparisons. These results suggest little benefit — to forecast quality — to generalizing our baseline model to allow departures from our MDS assumption. But they also show that doing so imposes little cost on forecast quality.

In our assessment, the ambiguity around the best model can be seen as consistent with the broader literature on survey forecasts. The strong evidence of bias and other departures from rationality in survey forecasts comes from in-sample analysis of the forecasts. Various studies have found these departures more challenging to exploit on an out-of-sample basis. Croushore (2010) documents that deviations of the SPF (and the Livingston Survey) forecasts from rationality are typically short-lived (episodic) and hard to exploit in real time; more recently, Foerster and Matthes (2022) and Hajdini and Kurmann (2024) have found similar results. Similarly, Eva and Winkler (2023) argue that the departures from rationality in Coibion and Gorodnichenko (2015) and subsequent studies do not apply on an entirely out-of-sample basis and cannot be used (in linear models) to improve on the SPF forecasts in real time. Biases being short-lived, contributing at most only a small fraction to overall MSE, is consistent with notions of SPF forecasts providing at least boundedly optimal, if not highly rational, forecasts, as discussed in, for example, Mertens and Nason (2020). On the other hand, Bianchi, Ludvigson, and Ma (2022) find that machine learning algorithms — fed large amounts of information — can be used to improve forecast accuracy in survey forecasts subject to time-varying bias. Ultimately, what model would we recommend? Our own inclination is for the simpler MDS model that yields forecasts that are fully SPF-consistent. But researchers with a stronger preference to allow for biases and other departures from rationality in SPF forecasts can comfortably proceed with the VAR specification.

5.3 SPF forecasts as compared to SEP projections

In the introduction, we illustrated the use of our approach to produce SEP-like fan charts for SPF-consistent forecasts. Of course, this raises a question as to how our SPF forecasts compare to SEP forecasts over a longer time period. As context, note that Reifschneider and Tulip (2019) showed that survey forecasts and Federal Reserve forecasts are very similar in terms of RMSE accuracy over long periods of time. Their comparison relied on the fixed-event annual forecasts published by various professional sources, including surveys; we will instead use our term structure of quarterly SPF forecasts to construct the fixed-event forecasts defined as in the SEP, which are Q4/Q4 growth and inflation rates and the Q4 unemployment rate. Since SEP forecasts are not available before late 2007, the sample for comparison is less than 20 years; so we rely on informal comparisons rather than engaging in formal inference that would likely be imprecise in the small sample available.[15]

To compare SEP forecasts with our forecasts constructed from the SPF using the MDS model, Figure 6 reports the errors in real-time forecasts (errors measured as realizations less forecasts) from the SEP and SPF, along with 68 percent uncertainty bands. The figure provides results for forecasts of the current year and the following two years. Note that, for each year from 2008 through 2023, we report estimates for SPF and SEP forecasts published every quarter; within each year, the (fixed-event annual) forecast horizon shrinks before resetting the next year, resulting in sawtooth patterns in the charts. In keeping with the SEP fan charts, in the SEP results in our figure the errors are based on point predictions measured as the median forecasts across participants, and the uncertainty bands are $+/-$ one times the historical RMSEs published in the SEP's Table 2. In the SPF results in the figure, the forecast errors are based on the posterior mean forecast obtained from our MDS model, and the uncertainty bands are based on the 16th and 84th percentiles of the predictive distribution.

The forecast errors reported in dots in the charts are broadly similar for the SEP and SPF, indicating that the forecasts were also broadly similar. Both the SEP and SPF-consistent projections of GDP growth, unemployment, and inflation made relatively large errors around the Great Recession and the pandemic. In some instances, such as with inflation following the pandemic, SPF forecasts were less accurate than SEP projections, whereas the reverse is true in other cases.

[Figure 6 about here.]

The 68 percent forecast uncertainty bands show more differences across the SEP and SPF than do the point forecasts. The SPF-consistent uncertainty bands from our MDS model are noticeably narrower than the SEP bands based on a simple 20-year rolling window of historical RMSEs of various professional forecasts. These differences are starker for GDP growth and unemployment than for inflation. The errors in SPF-consistent forecasts fall outside the SPF bands with some regularity, whereas the errors in SEP forecasts are rarely outside the SEP bands.[16] (Ideally, for correct coverage, about 32 percent of errors would be outside the bands.) The SEP appears to instead have bands wide enough to imply empirical coverage rates well above the nominal 68 percent rate. In this context, while the SPF-consistent bands are imperfect, in practice they could be a useful point of comparison to bands published in SEP fan charts.

Our SPF-consistent forecast uncertainty bands also differ from the SEP bands in their time variation. As might be expected, our estimates show more variation over time than do the SEP measures, because our models include stochastic volatility, whereas the SEP measures treat forecast error variances as being constant over 20-year windows. Following the Great Recession, our estimates of uncertainty rise notably; the SEP measures show much less change. In the aftermath of the recession, the SEP bands eventually widen a little, but they do so too late and too little in comparison to our model's estimates. In the ensuing economic recovery, our estimates fall quickly. From 2015 to 2019, both our estimates of uncertainty and the associated forecast error bands and the SEP estimates of uncertainty and forecast bands are relatively stable, with SEP bands wider than the SPF bands. Following the volatility induced by the outbreak of the pandemic in 2020, for GDP growth and unemployment our SV-based estimates of uncertainty rose sharply, to exceed the SEP estimates, resulting in much wider 68 percent forecast error bands, but only temporarily. For CPI inflation, our SPF-based estimates of forecast uncertainty have contours similar to those of the SEP bands until the pandemic. The pandemic's forecast errors initially caused uncertainty bands from our model to widen and then narrow, whereas the SEP bands trended upward with some delay.

6 Conclusion

This paper develops models that take published survey point forecasts — in our applications, from the SPF — as inputs and produce estimates of a longer, more complete term structure (across horizons) of survey-consistent point forecasts, along with a term structure of forecast uncertainty. Our estimates of forecast uncertainty reflect dispersion in past errors from the SPF's point forecasts, with time-varying uncertainty captured through stochastic volatility. Our methods can exactly replicate the SPF's quarterly forecasts at short horizons while interpolating through the available annual forecasts at longer horizons. To avoid excessively volatile imputation that can arise from small inconsistencies between quarterly and annual SPF predictions, we model annual forecasts with fat-tailed measurement errors that are a priori close to zero. Extending previous work, our models (1) provide SPF-consistent quarterly term structures of expectations and uncertainty that extend arbitrarily far ahead, while (2) handling the ragged edge of the SPF's fixed-event forecasts at longer horizons, and (3) offering a way to capture potential inefficiencies in SPF forecasts.

After illustrating the use of our approach to produce fan charts of annual forecasts (of Q4/Q4 percent changes or Q4 levels) directly analogous to those published by the FOMC, we show that our baseline model yields SPF-consistent quarterly real-time forecasts of GDP growth, unemployment, and inflation that match the published fixed-horizon quarterly forecasts from the SPF and interpolate through the annual fixed-event point forecasts. The model's estimates of forecast uncertainty vary over time, temporarily rising around recessions, and generally rise with the forecast horizon. The SPF forecasts obtained from our models perform comparably — and reasonably well — in various metrics of forecast quality, including efficiency as forecasts of the future SPF, unconditional coverage rates, and overall density calibration. The quality of our SPF-consistent estimated forecasts is comparable to the quality of the published short-horizon forecasts that have been widely used in research and practice. Our models generate fan charts with time-varying forecast uncertainty and reliable coverage rates.

Notes

[1]As surveyed in Clark and Mertens (2023), many studies have found that allowing for time-varying conditional variances improves the fit and forecasting performance of time series models.

[2]For examples of matching survey data with relatively coarse time series models, see Aruoba (2020), Crump, et al. (2023), Kozicki and Tinsley (2012), and Patton and Timmermann (2011).

[3]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 some of the sample.

[4]That is, we use the quarterly vintage in $t + h + 2$ to evaluate forecasts for $t + h$ made in t; this is the second estimate available in the RTDSM's vintages.

[5]Some previous studies have also made use of expectational updates, for different purposes. For example, Patton and Timmermann (2012) write a short-horizon forecast as a sum of a long-horizon forecast and forecast revisions, and use it as the basis of an optimal revision regression to test forecast optimality. Coibion and Gorodnichenko (2015) map out the implications of different theories of imperfect information for serial correlation in forecast updates.

[6]Other studies that decompose SPF forecasts into (perceptions of) permanent and transitory components include Clements (2022) and Krane (2011).

[7]An earlier working paper version of this paper derived an equivalent recursion for $\boldsymbol{Y}_t$; reflecting the common trend in $\boldsymbol{Y}_t$, this equivalent representation features a transition matrix with a single unit root and all other eigenvalues equal to zero.

[8]Our choice of a lag order of 1 is consistent with specifications and related evidence in Coibion and Gorodnichenko (2015) and CMM. In addition, given the limited amount of SPF bias reported in the prior literature (Section 2) , we view a low lag order as appropriate to limit the scope for overfitting.

[9]Mariano and Murasawa (2003) first developed the approximation in a nowcasting setting. Models of multi-period survey forecasts that rely on this (or similar) approximations include Aruoba (2020), Crump, et al. (2023), and Patton and Timmermann (2012).

[10]For CPI inflation, the mapping from y_t to $\bar{y}_t$ holds only approximately as well. For CPI, our definition of y_t as the simple percent change follows the SPF convention for its quarterly fixed-horizon forecasts, whereas the annual forecasts are solicited in terms of changes in Q4/Q4 levels.

[11]Since the SPF provides fixed-horizon forecasts for up to four quarters ahead, we disregard all current-year (fixed-event) forecasts. In Q4, there is also complete overlap between the SPF's annual forecast for next year and the observed fixed-horizon forecast, and we treat these observations with measurement error.

[12]When the forecast origin is in Q1, $h = 3$ points to the last quarter of the current year. For forecast origins later in the year, $h = 3$ points to the next year.

[13]The SPF generally provides fixed-horizon forecasts for quarters $h = 0$ through 4. Since the right-hand side of equation (20) requires observations for an $(h+1)$-step-ahead forecast, it is typical to pool forecasts for $h = 0$ through 3, which we do as well.

[14]Instances of relative RMSE that are worse for the VAR model compared to the MDS case are consistent with results reported in Bianchi, Ludvigson, and Ma (2022), who find generally poor out-of-sample predictability of SPF forecasts based on past forecasts alone.

[15]The available time series for the SEP begins in October 2007, and the charts lack SEP observations for 2020 because, in March 2020, in the aftermath of the pandemic's outbreak, the FOMC did not publish an SEP. For inflation, we compare SEP forecasts for PCE inflation with our CPI forecasts, since the SPF provides forecasts for PCE inflation only since 2007.

[16]In results omitted in the interest of brevity, we have verified that the patterns in results shown above for coverage rates of quarterly forecasts carry over to SEP-styled fixed-event forecasts.

References

Ang, Andrew, Geert Bekaert, and Min Wei (2007), "Do macro variables, asset markets, or surveys forecast inflation better?" *Journal of Monetary Economics*, 54, 1163–1212, `https://doi.org/10.1016/j.jmoneco.2006.04.006`.

Arias, Jonas E., Juan F. Rubio-Ramirez, and Minchul Shin (2023), "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Journal of Econometrics*, 235, 1054–1086, `https://doi.org/10.1016/j.jeconom.2022.04.013`.

Aruoba, S. Boragan (2020), "Term structures of inflation expectations and real interest rates," *Journal of Business & Economic Statistics*, 38, 542–553, `https://doi.org/10.1080/07350015.2018.1529599`.

Banbura, Marta, Federica Brenna, Joan Paredes, and Francesco Ravazzolo (2021), "Combining Bayesian VARs with survey density forecasts: Does it pay off?" Working Paper Series 2543, European Central Bank.

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma (2022), "Belief distortions and macroeconomic fluctuations," *American Economic Review*, 112, 2269–2315, `https://doi.org/10.1257/aer.20201713`.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2022), "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, forthcoming, `https://doi.org/10.1162/rest_a_01213`.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010), "The horseshoe estimator for sparse signals," *Biometrika*, 97, 465–480, `https://doi.org/10.1093/biomet/asq017`.

Chan, Joshua C. C. (2020), "Large Bayesian VARs: A flexible Kronecker error covariance structure," *Journal of Business & Economic Statistics*, 38, 68–79, https://doi.org/10.1080/07350015.2018.1451336.

Clark, Todd E., Michael W. McCracken, and Elmar Mertens (2020), "Modeling time-varying uncertainty of multiple-horizon forecast errors," *The Review of Economics and Statistics*, 102, 17–33, https://doi.org/10.1162/rest_a_00809.

Clark, Todd E., and Elmar Mertens (2023), "Stochastic volatility in Bayesian vector autoregressions," in *Oxford Research Encyclopedia in Economics and Finance*: Oxford University Press, https://doi.org/10.1093/acrefore/9780190625979.013.919.

Clements, Michael P. (2022), "Individual forecaster perceptions of the persistence of shocks to GDP," *Journal of Applied Econometrics*, 37, 640–656, https://doi.org/10.1002/jae.2884.

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, https://doi.org/10.1257/aer.20110306.

Croushore, Dean (2010), "An evaluation of inflation forecasts from surveys using real-time data," *The B.E. Journal of Macroeconomics*, 10, 1–32, https://doi.org/10.2202/1935-1690.1677.

Crump, Richard K., Stefano Eusepi, Emanuel Moench, and Bruce Preston (2023), "The term structure of expectations," in *Handbook of Economic Expectations* eds. by Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw: Academic Press, chap. 17, 507–540, https://doi.org/10.1016/B978-0-12-822927-9.00025-2.

Diebold, Francis X., and Roberto S. Mariano (1995), "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263, https://doi.org/10.2307/1392185.

Doh, Taeyoung, and A. Lee Smith (2022), "A new approach to integrating expectations into VAR models," *Journal of Monetary Economics*, 132, 24–43, https://doi.org/10.1016/j.jmoneco.2022.08.001.

Eva, Kenneth, and Fabian Winkler (2023), "A comprehensive empirical evaluation of biases in expectation formation," Finance and Economics Discussion Series 2023-042, Board of Governors of the Federal Reserve System (U.S.), https://doi.org/10.17016/FEDS.2023.042.

Farmer, Leland, Emi Nakamura, and Jon Steinsson (2024), "Learning about the long run," *Journal of Political Economy*, forthcoming, https://doi.org/10.1086/730207.

Faust, Jon, and Jonathan H. Wright (2009), "Comparing Greenbook and reduced form forecasts using a large realtime dataset," *Journal of Business & Economic Statistics*, 27, 468–479, https://doi.org/10.1198/jbes.2009.07214.

Foerster, Andrew, and Christian Matthes (2022), "Learning about regime change," *International Economic Review*, 63, 1829–1859, https://doi.org/10.1111/iere.12585.

Frey, Christoph, and Frieder Mokinski (2016), "Forecasting with Bayesian vector autoregressions estimated using professional forecasts," *Journal of Applied Econometrics*, 31, 1083–1099, https://doi.org/https://doi.org/10.1002/jae.2483.

Ganics, Gergely, and Florens Odendahl (2021), "Bayesian VAR forecasts, survey information, and structural change in the euro area," *International Journal of Forecasting*, 37, 971–999, https://doi.org/10.1016/j.ijforecast.2020.11.001.

Grishchenko, Olesya, Sarah Mouabbi, and Jean-Paul Renne (2019), "Measuring inflation anchoring and uncertainty: A U.S. and euro area comparison," *Journal of Money, Credit, and Banking*, 51, 1053–1096, https://doi.org/10.1111/jmcb.12622.

Hajdini, Ina, and André Kurmann (2024), "Predictable forecast errors in full-information rational expectations models with regime shifts," Working Paper 24-08, Federal Reserve Bank of Cleveland, https://doi.org/10.26509/frbc-wp-202408.

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, https://doi.org/10.1016/j.jeconom.2003.09.001.

Kim, Sangjoon, Neil Shephard, and Siddhartha Chib (1998), "Stochastic volatility: Likelihood inference and comparison with ARCH models," *The Review of Economic Studies*, 65, 361–393, https://doi.org/10.1111/1467-937X.00050.

Kozicki, Sharon, and P.A. Tinsley (2012), "Effective use of survey information in estimating the evolution of expected inflation," *Journal of Money, Credit, and Banking*, 44, 145–169, https://doi.org/10.1111/j.1538-4616.2011.00471.x.

Krane, Spencer D. (2011), "Professional forecasters' view of permanent and transitory shocks to GDP," *American Economic Journal: Macroeconomics*, 3, 184–211, https://doi.org/10.1257/mac.3.1.184.

Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo (2017), "Using entropic tilting to combine BVAR forecasts with external nowcasts," *Journal of Business & Economic Statistics*, 35, 470–485, https://doi.org/10.1080/07350015.2015.1087856.

Makalic, Enes, and Daniel F. Schmidt (2016), "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23, 179–182, https://doi.org/10.1109/LSP.2015.2503725.

Mariano, Roberto S., and Yasutomo Murasawa (2003), "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443, https://doi.org/10.1002/jae.695.

Mertens, Elmar (2016), “Measuring the level and uncertainty of trend inflation,” *The Review of Economics and Statistics*, 98, 950–967, `https://doi.org/10.1162/REST_a_00549`.

——— (2023), “Precision-based sampling for state space models that have no measurement error,” *Journal of Economic Dynamics and Control*, 154, p. 104720, `https://doi.org/10.1016/j.jedc.2023.104720`.

Mertens, Elmar, and James M. Nason (2020), “Inflation and professional forecast dynamics: An evaluation of stickiness, persistence, and volatility,” *Quantitative Economics*, 11, 1485–1520, `https://doi.org/10.3982/QE980`.

Mincer, Jacob A., and Victor Zarnowitz (1969), “The evaluation of economic forecasts,” in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance* ed. by Jacob A. Mincer: NBER, chap. 1, 3–46.

Newey, W.K., and K.D. West (1987), “A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica*, 55, 703–708, `https://doi.org/10.2307/1913610`.

Nordhaus, William D (1987), “Forecasting efficiency: Concepts and applications,” *The Review of Economics and Statistics*, 69, 667–674, `https://doi.org/10.2307/1935962`.

Patton, Andrew J., and Allan Timmermann (2011), “Predictability of output growth and inflation: A multi-horizon survey approach,” *Journal of Business & Economic Statistics*, 29, 397–410, `https://doi.org/10.1198/jbes.2010.08347`.

——— (2012), “Forecast rationality tests based on multi-horizon bounds,” *Journal of Business & Economic Statistics*, 30, 1–17, `https://doi.org/10.1080/07350015.2012.634337`.

Reifschneider, David, and Peter Tulip (2019), “Gauging the uncertainty of the economic outlook using historical forecasting errors: The Federal Reserve’s approach,” *International Journal* *of Forecasting*, 35, 1564–1582, https://doi.org/10.1016/j.ijforecast.2018.07.016.

Rossi, Barbara, and Tatevik Sekhposyan (2019), "Alternative tests for correct specification of conditional predictive densities," *Journal of Econometrics*, 208, 638–657, https://doi.org/10.1016/j.jeconom.2018.07.008.

Wright, Jonathan H. (2013), "Evaluating real-time VAR forecasts with an informative democratic prior," *Journal of Applied Econometrics*, 28, 762–776, https://doi.org/10.1002/jae.2268.

Table 1: Predictability of SPF point forecasts

	RGDP		UNRATE		PGDP		CPI	
Forecast	MDS	VAR	MDS	VAR	MDS	VAR	MDS	VAR
h = 0	1.41	1.32	0.86	0.83	0.99	0.98	1.15	1.19
	(0.24)	(0.25)	(0.08)	(0.10)	(0.05)	(0.06)	(0.16)	(0.15)
h = 1	1.02	1.02	0.91	0.86	1.02	0.99	1.01	1.00
	(0.09)	(0.09)	(0.06)	(0.08)	(0.04)	(0.04)	(0.07)	(0.06)
h = 2	1.02	0.97	0.94	0.90	**0.94**	**0.85**	0.94	0.95
	(0.08)	(0.08)	(0.06)	(0.07)	(0.03)	(0.03)	(0.04)	(0.05)
h = 3	0.94	**0.80**	0.96	0.91	**0.92**	**0.86**	**0.92**	**0.92**
	(0.09)	(0.09)	(0.05)	(0.06)	(0.04)	(0.03)	(0.04)	(0.04)
h = 4	**0.87**	**0.57**	0.97	0.93	**0.91**	**0.90**	0.94	**0.89**
	(0.06)	(0.09)	(0.05)	(0.06)	(0.03)	(0.04)	(0.04)	(0.03)
y = 1	0.94	0.91	0.96	0.93	**0.93**	**0.91**	0.98	0.96
	(0.09)	(0.07)	(0.05)	(0.06)	(0.03)	(0.03)	(0.05)	(0.05)
y = 2	0.94	0.96	0.92	0.95	—	—	0.85	**0.74**
	(0.09)	(0.10)	(0.07)	(0.04)	—	—	(0.11)	(0.09)
y = 3	0.95	0.59	**0.76**	0.95	—	—	—	—
	(0.06)	(0.24)	(0.08)	(0.04)	—	—	—	—

Notes: Estimated slope coefficients of Mincer-Zarnowitz regressions for model-based predictions of next-quarter's published values for SPF forecasts at different forecast horizons. Heteroskedasticity-consistent standard errors in parentheses. Bold font distinguishes coefficient estimates significantly different from 1 with a 10% confidence level. Evaluation window from 1990Q1 to 2023Q4 (and as far as data for SPF forecasts at the different horizons are available).

Table 2: Relative forecast accuracy of MDS vs. VAR model

	RMSE				CRPS			
h	RGDP	UNRATE	PGDP	CPI	RGDP	UNRATE	PGDP	CPI
0	1.01	1.12	0.99	0.91**	1.02	0.95	1.00	0.93***
1	1.04	1.05	1.02	1.01	1.01	1.00	1.00	1.01
2	0.99	1.07	1.01	1.01	0.99	1.01	0.99	1.01
3	1.00	1.04	1.02	1.01	1.01	1.00	1.00	1.01
4	1.00	1.02	1.03	1.01	1.00	1.00	1.00	1.01
5	1.00	1.02	1.04	1.01	1.00	1.01	1.01	1.01
6	1.00	1.02	1.04	1.01	1.01	1.02	1.02	1.01
7	1.00	1.02	1.04	1.01	0.99	1.03	1.02	1.01
8	1.00	1.02*	1.04	1.01	1.00	1.03	1.02	1.01
9	1.00	1.03*	1.04	1.01	1.00	1.04	1.03	1.01
10	1.00	1.03	1.04	1.01	1.00	1.04	1.03	1.02
11	1.00	1.02	1.04	1.01	1.00	1.03	1.03	1.02
12	1.00	1.01	1.04	1.01	1.00	1.02	1.03	1.01
13	1.00	1.00	1.04	1.01	1.00	1.00	1.03	1.01
14	1.00	0.99	1.04	1.01	1.01	0.99	1.04	1.01
15	1.00	1.00	1.04	1.01	1.01	1.00	1.03	1.01
16	1.00	1.00	1.04	1.00	1.01	0.99	1.03	1.01

Note: Relative RMSE and CRPS of VAR model (with MDS in denominator). Quarterly forecast horizons, h. Evaluation window from 1990Q1 through 2023Q4 (and as far as realized values are available). Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h + 1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

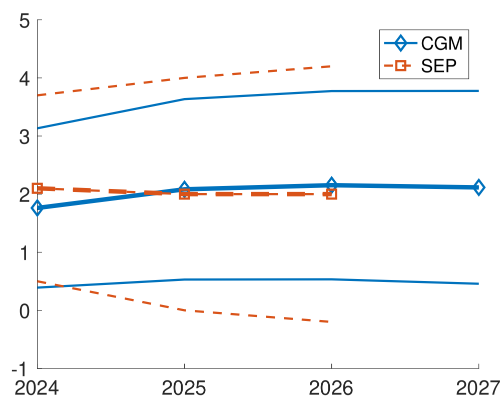
Table 3: Forecast coverage rates (VAR model)

	RGDP		UNRATE		PGDP		CPI	
h	68%	90%	68%	90%	68%	90%	68%	90%
0	50.00***	81.62***	76.47*	94.85**	55.88***	84.56*	66.18	91.18
1	55.56***	79.26***	77.78**	93.33	62.22	82.22**	60.74	84.44
2	58.21**	82.84**	77.61*	94.03	65.67	85.82	63.43	84.33
3	53.38***	82.71*	78.20*	91.73	64.66	86.47	64.66	86.47
4	56.06**	85.61	68.94	90.91	62.88	86.36	64.39	84.09
5	59.54	86.26	66.41	89.31	64.89	87.79	68.70	87.02
6	58.46**	86.15	63.85	88.46	64.62	87.69	66.15	88.46
7	64.34	86.05	62.02	88.37	64.34	89.92	65.89	87.60
8	62.50	84.38	60.16	87.50	65.62	88.28	68.75	86.72
9	62.20	83.46	56.69*	86.61	66.93	86.61	66.14	88.19
10	62.70	84.92	56.35*	85.71	65.87	89.68	67.46	87.30
11	66.40	83.20	52.00**	83.20	69.60	88.80	72.00	88.00
12	63.71	84.68	50.00**	82.26	69.35	89.52	70.16	87.90
13	69.11	84.55	48.78**	82.93	67.48	90.24	70.73	87.80
14	66.39	86.89	49.18**	82.79	69.67	89.34	69.67	88.52
15	64.46	86.78	48.76**	82.64	69.42	88.43	69.42	89.26
16	66.67	87.50	48.33*	81.67	71.67	90.00	71.67	89.17

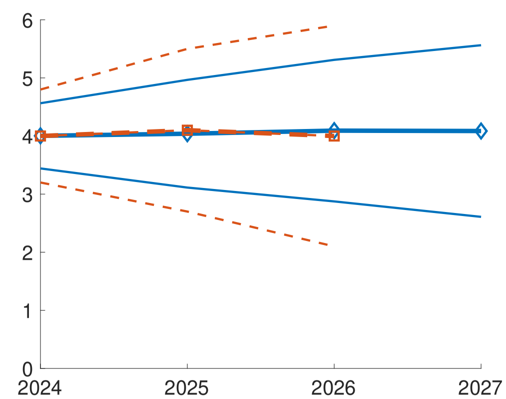
Note: Coverage rates for uncertainty bands with nominal levels of 68% and 90% for out-of-sample forecasts at quarterly forecast horizons, h. Evaluation window from 1990Q1 through 2023Q4 (and as far as realized values are available). Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h + 1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Figure 1: Fan charts and SEP (2024Q1)

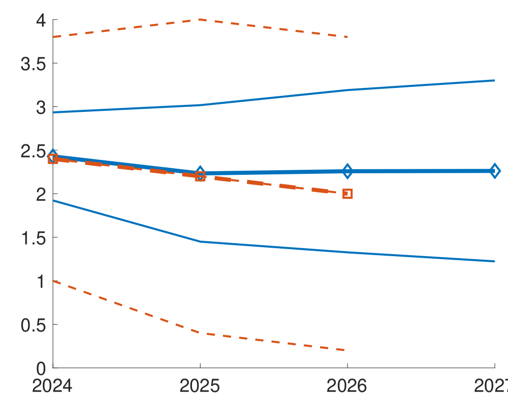
(a) Real GDP growth



(b) Unemployment rate

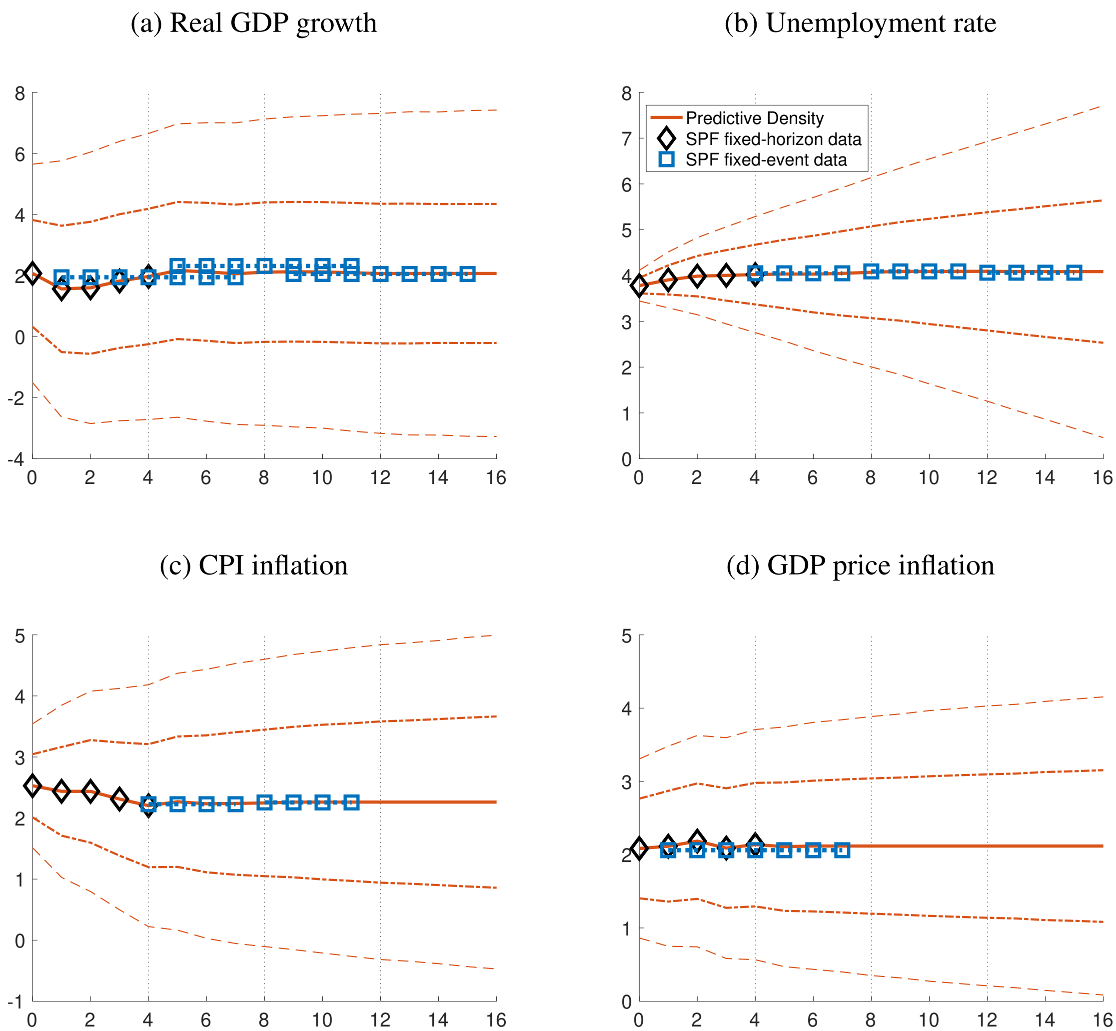


(c) Inflation



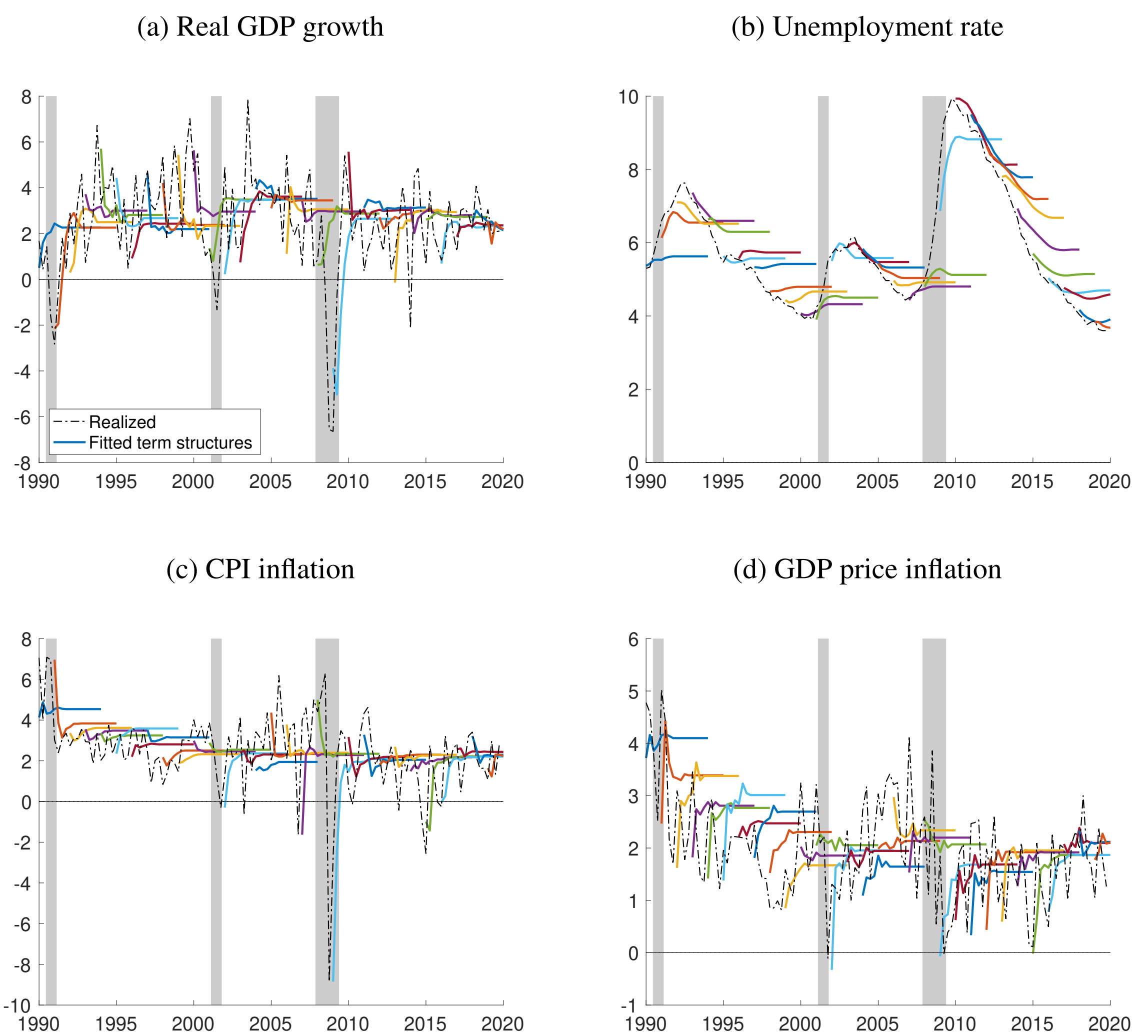
Notes: Model-based annual predictions, with 68% predictive intervals, from our model ("CGM") and the FOMC's SEP per 2024Q1. The predictions are calendar-year forecasts of Q4/Q4 growth rates of GDP and prices and the Q4 level of the unemployment rate. Since the forecast origin is early in the year, the SEP extends only two years out. For inflation, Panel (c) compares model-based densities for CPI against the SEP's forecasts for PCE inflation. The version of the model used to generate these figures treats SPF forecasts as martingales.

Figure 2: Quarterly fan charts and SPF per 2024Q1



Notes: SPF-consistent predictive densities as generated from our MDS model, showing predictive mean as well as 68% and 90% uncertainty bands. Quarterly forecast horizons on the horizontal axis. Diamonds indicate observed values for SPF fixed-horizon forecasts for $h = 0, 1, 2, 3, 4$. Squares depict observed values for fixed-event calendar-year predictions from the 2024Q1SPF. The fixed-event calendar-year predictions represent linear combinations of (implied) forecasts for multiple quarters, and squares are shown for all quarters involved, with dots connecting the squares associated with a given calendar year. (As explained in the text, annual SPF predictions for GDP growth and GDP price inflation involve quarters in adjacent years.)

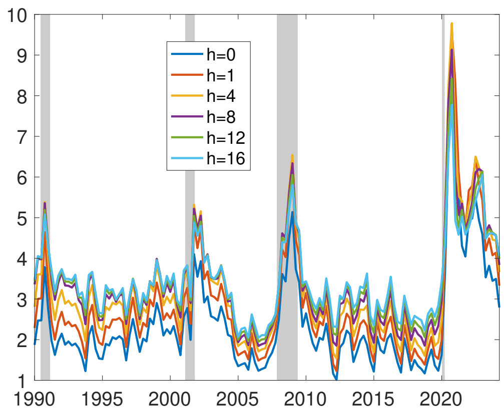
Figure 3: SPF-consistent term structures of expectations over time (MDS)



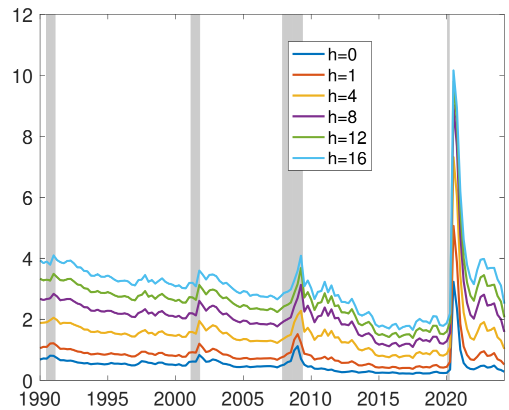
Note: Term structures of expectations for quarterly forecasts, generated out-of-sample from our MDS model at different forecast origins, and realized values. For sake of readability, only forecast origins in Q1 and for years prior to the COVID-19 pandemic are shown. Shaded areas depict NBER recessions.

Figure 4: Term structures of uncertainty over time (MDS)

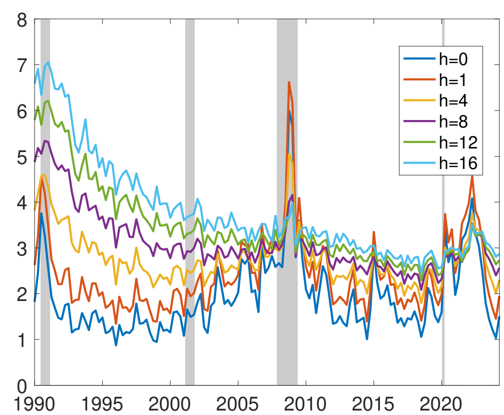
(a) Real GDP growth



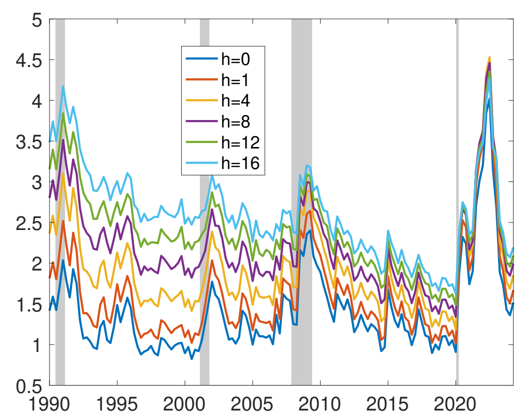
(b) Unemployment rate



(c) CPI inflation

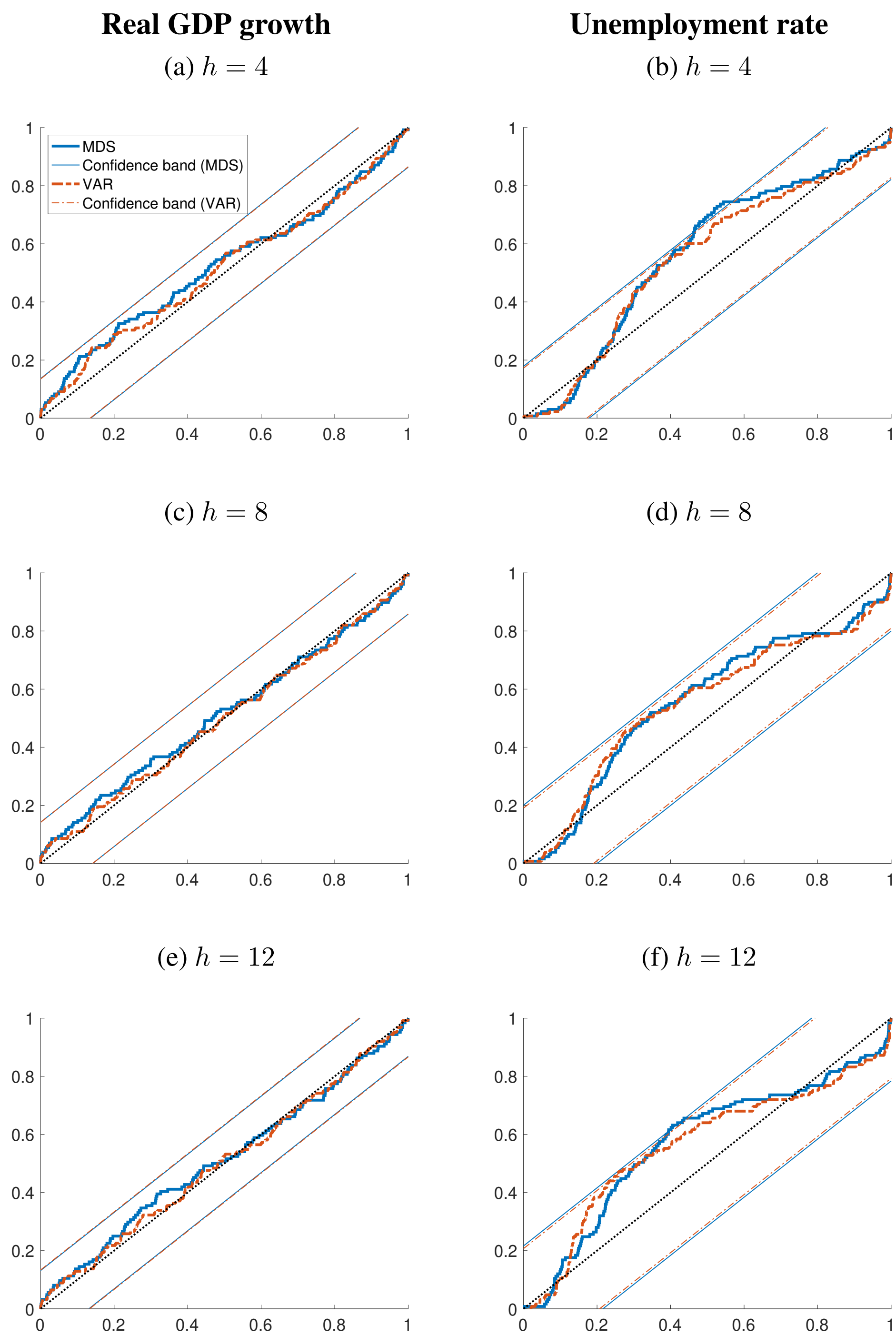


(d) GDP price inflation



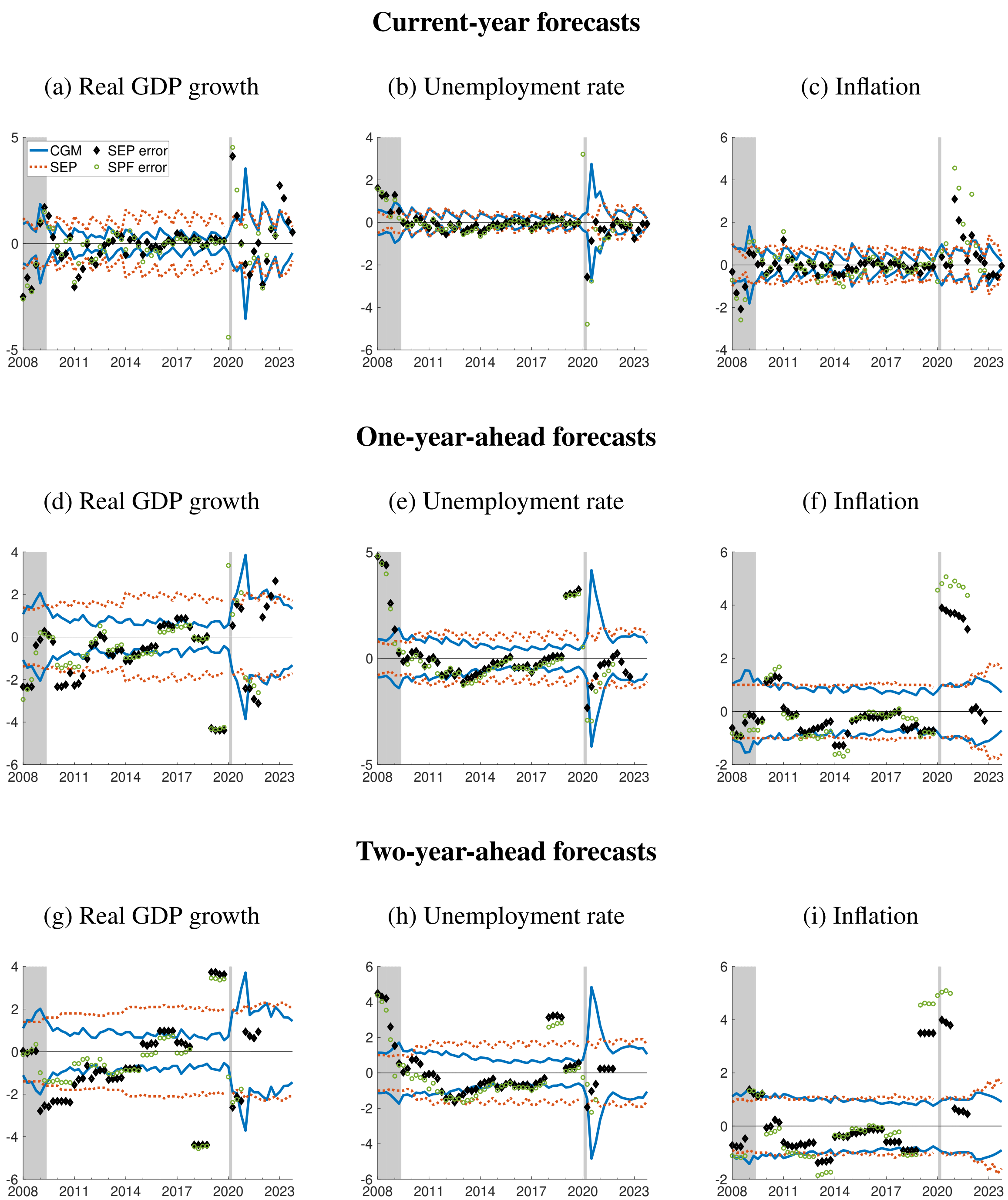
Note: Uncertainty measured by the width of 68% predictive intervals, generated out of sample from our MDS model, for selected quarterly forecast horizons. Shaded areas depict NBER recessions.

Figure 5: PITs for real GDP growth and the unemployment rate



Notes: Empirical cumulative distributions of probability integral transforms (PITs) for GDP growth and unemployment at selected quarterly forecast horizons. All forecasts are generated out of sample by our MDS and VAR models, and evaluated over an evaluation window from 1990Q1 through 2023Q4 (and as far as realized values are available). 95% confidence bands for tests of correct calibration from Rossi and Sekhposyan (2019); computed separately for each model, but with nearly identical plot lines.

Figure 6: Error bands of annual forecasts: MDS model vs. SEP



Note: Forecast errors measured as realizations less forecasts. Model-based uncertainty bands correspond to 68% predictive intervals. FOMC's SEP bands reflect the historical RMSEs of professional forecasts over the previous 20 years as described by Reifschneider and Tulip (2019). SPF-consistent densities and (ex-post) errors for calendar-year definitions of the SEP generated from our MDS model. For inflation, we compare model-based densities for CPI against the SEP's forecasts for PCE inflation.