Constructing the Term Structure of Uncertainty 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]*Central Bank of Hungary,* [3]*Deutsche Bundesbank.*

July 8, 2022

Abstract

We develop a model that permits the estimation of a term structure of both expectations and forecast uncertainty for application to professional forecasts such as the SPF. Our approach exactly replicates a given data set of predictions from the SPF or a similar forecast source without measurement error. Extending some previous work, our model includes not only fixed-horizon forecasts but also fixed-event forecasts, including at horizons beyond the fixed-horizon maximum, and it accommodates variation over time in the available horizons of fixed-event forecasts. The model casts a decomposition of multi-period forecast errors into a sequence of forecast updates that may be partially unobserved, resulting in a multivariate unobserved component model with stochastic volatility. In addition, we bring the density forecast information contained in the SPF's probability bins for fixed-event forecasts to bear on the term structure of uncertainty through entropic tilting. This application of entropic tilting allows us to treat the SPF's subjective point and density forecasts commensurately.

In our empirical analysis, examples of quarterly fan charts of point forecasts and uncertainty bands around those forecasts show that the model's estimates of uncertainty vary over time. Over the full sample, incorporating information in the SPF's subjective probability bins through entropic tilting does not improve anymore on the baseline model estimates, but in some subsamples, tilting to the bin information improves the accuracy of both point and density forecasts. We conclude by applying our estimates to construct SPF-based fan charts with calendar year forecasts like those published by the Federal Reserve.

Keywords: Term structure of expectations, uncertainty, survey forecasts, fan charts, entropic tilting
JEL classification codes: E37, C53

*We gratefully acknowledge helpful suggestions and comments received from Refet Gurkaynak, James Mitchell, and workshop or conference participants at the Deutsche Bundesbank, the 2022 Dolomiti Macro Meeting, Heidelberg University, and the 2022 IAAE conference. 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 Deutsche Bundesbank, the Central Bank of Hungary, or the Eurosystem.

1 Introduction

The macroeconomic projections of professional forecasters are widely used in both economic policymaking and forecasting research. Such forecasts with long histories include the (US) Survey of Professional Forecasters (SPF), Blue Chip Consensus, Consensus Economics, and IHS Markit (formerly Macroeconomic Advisers), as well as Federal Reserve forecasts published in the Tealbook or Greenbook and the Federal Open Market Committee's (FOMC) Summary of Economic Projections. These forecasts have a wealth of useful information, but with some unevenness in availability across forecast horizons. For example, the SPF includes both (1) fixed-horizon quarterly point forecasts, at shorter horizons, and (2) fixed-event annual forecasts, covering shorter and longer horizons. Although most professional forecasts focus on point predictions, one unique feature of the SPF compared to other forecast sources is that it provides fixed-event density forecasts in the form of probability bins.[1] For reasons that will become clearer below, our analysis focuses on the SPF in order to be able to make use of survey-based density forecasts. In any event, we characterize the available information as the "ragged edge of SPF forecasts."

A number of studies have developed models that use the mixtures of forecast horizons available in professional forecasts to build out a more complete term structure (across horizons) of point forecasts. For example, Kozicki and Tinsley (2012) and Aruoba (2020) develop models to extend point forecasts to obtain a more complete term structure of inflation forecasts. Focusing on capturing time-variation in long-run forecasts, Crump, et al. (2021) develop a multivariate unobserved components model of trend and cycle to fit a range of macroeconomic survey forecasts to estimate a term structure of point forecasts in growth, inflation, and a short-term interest rate. In a similar vein, Hepenstrick and Blunier (2022) develop an approach for interpolating a term structure of complete quarterly forecasts from available quarterly and annual forecasts using a state-space representation that includes a simple time series process for quarterly growth and growth forecasts and a measurement equation that relates the forecasts of interest to the available measurements. Some other work (e.g., Dovern, Fritsche, and Slacalek (2012)) has taken a simpler interpolation approach to translate fixed-event point forecasts to fixed-horizon predictions, and Ganics, Rossi, and Sekhposyan (2021) develop an approach for translating fixed-event density forecasts to fixed-horizon

[1] The available sample of forecasts from SPF is also longer than those of most other professional forecasts.

quarterly forecasts.

We develop a model that permits not only the estimation of a term structure of expectations, but also a term structure of forecast uncertainty. In addition, our methods allow us to exactly replicate a given data set of predictions (from the SPF or other judgmental forecast sources) without measurement error. Forecast uncertainty is widely recognized to be important for monetary policy decisions, and many central banks (e.g., the Bank of England and the FOMC) publish estimates of their forecast uncertainty as an integral part of their policy communications. Clark, McCracken, and Mertens (2020) (hereafter, CMM) develop a model that uses realized errors in SPF point forecasts to estimate forecast uncertainty, pooling information embedded in forecast errors for different forecast horizons. Echoing results from prior research on model-based forecasts — typically based on vector autoregressions (VARs) and dynamic stochastic general equilibrium (DSGE) models — CMM find that model fit and forecast performance are improved by allowing for time-varying conditional variances. Specifically, CMM employ a multivariate stochastic volatility specification that improves the accuracy of uncertainty measures for survey forecasts compared to simpler approaches for tracking variances.[2] While CMM can be seen as yielding a term structure of forecast uncertainty, the maximal horizon of their application is limited to the four-quarter fixed-horizon forecasts of SPF.

In this paper, to obtain a longer and more complete term structure of forecasts and their uncertainty, we extend CMM in three directions. First, we generalize the model to include not only fixed-horizon forecasts but also fixed-event forecasts, including at horizons beyond the fixed-horizon maximum of four quarters (ahead). Second, we also generalize the model to accommodate the ragged edge of SPF forecasts due to fixed-event horizons varying systematically within each year of the quarterly publication of the survey and due to the survey adding more years of fixed-event horizons some years ago. Together, these extensions permit us to use the ragged edge of SPF forecasts to obtain complete quarterly forecast fan charts at each forecast origin, extending to a horizon of 15 quarters.

Finally, we bring the density forecast information contained in the SPF's probability bins for

[2]Examples of VAR models with stochastic volatility include Clark (2011), D'Agostino, Gambetti, and Giannone (2013), and Clark and Ravazzolo (2015). Reifschneider and Tulip (2019) provide simple evidence of changes in the sizes of forecast errors associated with projections from the Federal Reserve and other sources, including the SPF and Blue Chip Consensus.

fixed-event forecasts to bear on the term structure of uncertainty through entropic tilting. Entropic tilting can be seen as a non-parametric approach to conditional forecasting, used in other studies such as Cogley, Morozov, and Sargent (2005), Ganics and Odendahl (2021), Krüger, Clark, and Ravazzolo (2017), Robertson, Tallman, and Whiteman (2005), and Tallman and Zaman (2020). In this application, we use tilting to adjust the model-based predictive distribution so that the tilted distribution's implied probability bins match up to those from the SPF (while minimizing a distance criterion). A further contribution of our paper is thus to target directly the bin probabilities, not a more limited set of specific moments, like means and variances. In contrast, others have instead used entropic tilting to match estimated values for mean and variances that were obtained from a history of observed SPF point forecasts (see, for example, Krüger, Clark, and Ravazzolo (2017)). Alternatively, other work, including Galvão, Garratt, and Mitchell (2021) and Grishchenko, Mouabbi, and Renne (2019), have instead targeted values for second or third moments that were imputed from the probability bin forecasts, by fitting density functions to the survey histograms.

One motivation for the use of entropic tilting is that it may help improve the uncertainty estimate and density forecast accuracy, particularly at horizons longer than the quarters covered in a given published SPF's quarterly projections. Another reason is that tilting allows us to treat the SPF's subjective point and density forecasts commensurately. In our baseline model, we take the SPF point forecasts as *the* forecasts and do not try to improve on them with a model; that is, while our approach can be seen as interpolating quarterly forecasts between annual fixed-event forecasts, we do not second-guess the observed SPF's point forecasts. Using entropic tilting to hit the SPF's subjective probability bins can be seen as also matching up the model's uncertainty estimates for the associated fixed-event forecasts to the SPF-implied estimates, such that the tilted model's estimates include no second-guessing. To the best of our knowledge, we are the first paper to incorporate the SPF's histogram bins directly into entropic tilting of model-based predictive densities.

In addition, we consider an extended version of our model that allows for bias in observed SPF forecasts, resulting in persistent forecast errors. This extended model treats SPF predictions as unbiased martingales only in its prior, while allowing for generic VAR dynamics of SPF forecasts in its posterior. However, when comparing the average forecast performance of point and density

forecasts over time, the extended model performs very closely to, and not clearly better than our baseline specification. These results are consistent with the existing literature that finds deviations from SPF forecast efficiency to be episodic and hard to consistently exploit in real time.[3]

In our empirical analysis, we focus on SPF forecasts of GDP growth — for which the SPF sample of annual forecasts is largest — and more briefly report results for unemployment and inflation, which are generally consistent with those for GDP growth. We first show that our model yields quarterly forecasts that perfectly match and interpolate through the annual fixed-event point forecasts; sometimes the interpolation is quite smooth, but in other instances, it shows some variation of quarterly forecasts around the annual predictions. We next provide examples of quarterly fan charts of point forecasts and uncertainty bands around those forecasts. These show that the model's estimates of uncertainty vary over time and generally rise with the forecast horizon.

Our third set of results examines the efficacy of using entropic tilting to incorporate information in the SPF's subjective probability bins for fixed-event annual forecasts. We provide examples to illustrate the impacts of bin-tilting on predictive distributions, compare the historical accuracy of the SPF's bins and model-based bins in annual forecasts, and compare bin-tilting's impacts on the historical accuracy of quarterly forecasts. These formal evaluations show that, over the full sample, the model's point and density forecasts based on SPF point forecasts are relatively good; on average, tilting to the SPF's probability bins does not offer consistent benefits. That said, there can be periods in which tilting to the bin information helps forecast accuracy — for both point and density forecasts — including in GDP growth over the period of 2009-2016 and unemployment in the period of the COVID-19 pandemic. A fourth set of results shows that the historical accuracy of point and density forecasts from the version of our model extended to allow for bias in observed SPF forecasts is generally very similar to the accuracy of our baseline model. So while on the one hand there is little benefit to generalizing our baseline model to allow for bias in SPF point forecasts, on the other there is little cost in forecast accuracy to doing so.

We conclude the paper with a practical application using our estimated term structure of forecasts: constructing fan charts with calendar year (four-quarter growth rates) forecasts like

[3]See, for example, Croushore (2010), or more recently, Bianchi, Ludvigson, and Ma (2022), Foerster and Matthes (2022), Hajdini and Kurmann (2022), and Mertens and Nason (2020).

those published by the FOMC. We show that, in recent years, our model applied to SPF forecasts yields estimates of uncertainty around GDP growth and unemployment forecasts that are lower than those implied by the historical forecast RMSEs that underlay the FOMC's fan charts. Our approach could be used in real time with each new release of the SPF to produce updated fan charts of point forecasts and estimates of forecast uncertainty.

The paper proceeds as follows. Section 2 provides an overview of related literature on survey forecasts not covered above. Section 3 describes the SPF forecasts and data used in the evaluation. Section 4 presents our model, briefly describes estimation, and details the entropic tilting we use to incorporate information from the SPF's probability bins. Section 5 provides results. Section 6 concludes.

2 Related Literature

A long literature — more than can be covered here — has 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. Coibion and Gorodnichenko (2015) develop a new approach to testing rational expectations that permits quantifying departures from full rationality and the degree of information rigidities; their applications include inflation forecasts from the SPF. Focusing on inflation forecasts, Ang, Bekaert, and Wei (2007) find survey forecasts hard to beat by a battery of forecasting methods, and Croushore (2010) documents that deviations of SPF (and Livingston survey) from rationality are typically short-lived and hard to exploit in real time. In this spirit, Mertens and Nason (2020) propose a new unobserved component model of inflation that distinguishes trend and inflation gap components and features a sticky information forecast mechanism; their estimates reveal that the stickiness of survey forecasts is not invariant to the time series process governing actual inflation. Using machine learning algorithms, Bianchi, Ludvigson, and Ma (2022) find evidence of time-varying bias in survey expectations and forecasts and conclude that artificial intelligence can be used to improve forecast accuracy. Regarding the predictive value of density forecasts collected by surveys, Clements (2018), as well as Glas

and Hartmann (2021) and others, point to potential shortcomings, for example due to rounding of answers by respondents.

Another long literature has sought to use professional forecasts to improve forecasts from time series models. In one example, 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 Bayesian VAR forecast accuracy can be improved by using long-run survey forecasts as priors on the long-run means of the model. Krüger, Clark, and Ravazzolo (2017) improve forecasts from Bayesian VARs through entropic tilting toward nowcasts from survey-based forecasts. Frey and Mokinski (2016) instead improve forecasts from Bayesian VARs by adding survey-based nowcasts as endogenous variables in the VAR and using priors so that the dynamics of the survey forecasts inform the parameter estimates of the dynamics of the actual data.[4]

Many other studies (in addition to those noted above) have used professional forecasts to examine the term structure of forecast uncertainty across horizons or time variation in uncertainty at given horizons. 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-squared forecast errors. Clements and Galvão (2017) compare ex ante uncertainty estimates and ex post root-mean-squared errors (RMSEs) from annual fixed-event SPF forecasts as well as corresponding measures from time series models of growth and inflation. To capture and assess time variation in macroeconomic uncertainty, Jo and Sekkel (2019) develop and estimate a factor stochastic volatility model using errors in SPF point forecasts of a small set of macroeconomic variables.

Yet another long line of work has sought to develop ways to use the SPF's subjective, fixed-event probability bins for more general density forecasting purposes. Many have sought — see Clements and Galvão (2017) and references therein — to compare model-based probability densities (or moments thereof) to the bin forecasts. Ganics, Rossi, and Sekhposyan (2021) develop a method for translating fixed-event density forecasts from the SPF to fixed-horizon quarterly forecasts. They rely on density combination that weights the fixed-event density forecasts according to uniformity of the probability integral transform criterion, aiming at obtaining a correctly calibrated fixed-

[4]In a similar vein, Doh and Smith (2021) develop priors that align a VAR's (a priori) forecasts with survey predictions.

horizon density forecast. Recently, Bassetti, Casarin, and Del Negro (2022) develop a Bayesian non-parametric approach to density estimation from the bin probabilities.

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. Reflecting in part the forecasts available, we examine quarterly and annual forecasts from the SPF for a basic set of major macroeconomic aggregates: GDP growth (RGDP), the unemployment rate (UNRATE), and inflation in the GDP price index (PGDP).[5] (For simplicity, we use "GDP" and "GDP price index" to refer to output and price series, even though, in our real-time data, the measures are based on GNP and a fixed-weight deflator for much of the sample.) These variables are commonly included in research on the forecasting performance of models such as VARs or DSGE models. The various forecast sources analyzed by Reifschneider and Tulip (2019) cover a very similar set of variables.[6] The SPF forecasts are widely studied, publicly available, and the longest available time series of forecasts for a range of variables. Alternatives such as the Blue Chip Consensus are not available publicly or for as long a sample, and they lack the density forecast information contained in the SPF.

[Table 1 about here.]

We obtained the SPF forecasts of growth, unemployment, and inflation from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM). Reflecting the data available, our estimation samples start with 1969Q1, and the sample end point is 2022Q2. 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 (with some exceptions noted below). In

[5]The SPF defines the relevant unemployment rate measure as the quarterly average of monthly data. For real GDP and its price index, the SPF solicits point forecasts in terms of levels, which we convert into growth rates. In order to map calendar-year predictions of growth rates pertaining to changes in yearly average levels into out model, we employ a log-linear approximation detailed in Section 4.3. Accordingly, we convert SPF point forecasts for real GDP and its price index into continuously compounded growth rates.

[6]Reflecting the availability of histogram data, the set of variables covered is a subset of the five variables covered by Clark, McCracken, and Mertens (2020), who also considered SPF point forecasts for CPI and the T-bill rate. The state space model developed in our paper can, of course, also be applied to those variables. In results not included here (for brevity) we found the model to also work well when applied to CPI and the T-bill rate.

2009Q2, the forecast horizon for GDP growth and unemployment was extended to include annual forecasts for two additional years, i.e., two and three years ahead. Table 1 lists the first available dates for SPF forecasts of different variables at different horizons.

The availability of forecasts for probability bins also evolved over time, with some differences with respect to the point forecasts. Beyond a current-year horizon, the SPF only provides probability bins for GDP growth, the unemployment rate, and inflation in the GDP price index.[7] The sample of probability bins for unemployment begin with the 2009Q2 SPF. In the case of GDP and the GDP price index, we use bin data beginning with 1992Q1. Although the SPF provides probability bins for these variables from 1981 through 1991, these earlier years pose a number of what Diebold, Tay, and Wallis (1999) refer to as "complications." These complications include some shifts in the number of bins and their ranges and changes in forecast periods — including, as noted in the Philadelphia Fed's SPF documentation, some uncertainty as to the horizons of the annual forecasts covered in the 1985Q1 and 1986Q1 surveys.[8] To avoid possible distortions from these complications, we only use probability bin forecasts starting with 1992 in our analysis with entropic tilting.

In all cases, we form the point forecasts using the average over all SPF responses. For the fixed-event probability bin forecasts we also use average SPF responses. The average probability predictions can be seen as forecasts that would be obtained with linear pooling of the underlying probability forecasts of individual participants of the SPF. As summarized in such sources as Bassetti, Casarin, and Del Negro (2022), simple linear pooling of density forecasts has worked well in other settings.

As an example of the available forecast horizons and their variation within the calendar year, the 2021Q4 SPF reported point forecasts of GDP growth and the unemployment rate for fixed quarterly horizons of the current and next four quarters, as well as fixed-event annual forecasts for the current year and the next three. Within a year of SPF publications, the effective maximum of implied quarterly horizons varies across quarters. In this fourth-quarter example, the last annual forecast extends 12 quarters ahead (the annual forecast reported for 2024 includes 2024Q4, 12

[7] The SPF began to report bins — along with point forecasts — for core CPI and core PCE inflation in 2007. We omit these measures from our analysis due to the short sample of available forecasts.

[8] See p.37 of the documentation available at `https://www.philadelphiafed.org/-/media/frbp/assets/surveys-and-data/survey-of-professional-forecasters/spf-documentation.pdf`.

quarters beyond the 2021Q4 forecast origin). In the 2022Q1 SPF, the last annual forecast for 2025 includes 2025Q4, 15 quarters beyond the forecast origin. For probability bin forecasts, the 2021Q4 SPF reported bin forecasts of GDP growth and the unemployment rate for fixed-event horizons of annual forecasts for the current and subsequent three years (as noted above, the SPF reports probability forecasts for calendar-year forecasts and not the quarterly forecasts).

For growth and inflation rates at an annual frequency, the SPF uses certain conventions we incorporate in the measurement specification of our model, as detailed in Section 4. Whereas all quarterly forecasts of growth and inflation use annualized quarter-on-quarter rates, the annual forecasts of GDP growth and inflation in the GDP price index refer to percent changes in annual average levels of GDP and the price index.

To estimate our model, we also need measures of the outcomes of the variables. In the case of GDP growth and GDP inflation, data can be substantially revised over time. Specifically, for GDP growth and GDP inflation, we obtain real-time measures for quarter $t-1$ data as these data were publicly available in quarter t from the quarterly files of real-time data in the RTDSM. As described in Croushore and Stark (2001), the vintages of the RTDSM are dated to reflect the information available around the middle of each quarter. For forecast evaluation, we use the second-available estimate from the RTDSM to measure the outcomes of GDP growth and GDP inflation (that is, we use the quarterly vintage in $t+h+2$ to evaluate forecasts for $t+h$ made in t). Because revisions to quarterly data are relatively small for the unemployment rate, 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 this variable.

4 Model and Estimation

We begin by specifying the general form of our baseline model and then proceed to explain its pieces and complexities in more detail. In broad terms, our model can be seen as a multivariate unobserved component model with stochastic volatility; we express it in state-space form with stochastic volatility in the state equation. We design the state-space specification to match arbitrary term structures of expectations, with application to the SPF in this paper.

In all cases, we specify and estimate the model on a variable-by-variable basis (i.e., we estimate

it separately for GDP growth, the unemployment rate, and inflation). The variable of interest in quarter t is denoted y_t, and forecasts for period $t+h$ from forecast origin t are denoted $y_{t+h|t}$. Let $\boldsymbol{Y}_t$ denote a (partially latent) state vector containing the lagged realized value and a term structure of quarterly fixed horizon forecasts. Specifically, $\boldsymbol{Y}_t$ consists of the lagged realized value y_{t-1} (that is observed at t), the time t nowcast and quarterly fixed-horizon forecasts extending from $h = 1$ up to the maximum quarterly horizon that can be covered in the historical SPF data. We denote this maximal horizons by H and obtain: $\boldsymbol{Y}_t = \left(y_{t-1}, y_{t|t}, y_{t+1|t}, \ldots, y_{t+H|t}\right)'$. Some of the elements of $\boldsymbol{Y}_t$ are unobserved. The available measures of SPF point forecasts — for quarterly horizons up to 4 steps ahead and for annual fixed events up to 4 years ahead — are collected in the measurement vector $\boldsymbol{Z}_t$.

4.1 An accounting identity for forecasts

We build our model from an accounting identity, also used by CMM, that decomposes h-step-ahead forecast errors into the sum of the $t+h$ nowcast error, as well as preceding forecast updates:

$$y_{t+h} - y_{t+h|t} = e_{t+h} + \sum_{i=1}^{h} \mu_{t+h|t+i}, \tag{1}$$

$$\text{with} \qquad e_{t+h} \equiv y_{t+h} - y_{t+h|t+h}\,, \tag{2}$$

$$\text{and} \qquad \mu_{t+h|t+i} \equiv y_{t+h|t+i} - y_{t+h|t+i-1}\,, \tag{3}$$

so that e_{t+h} denotes the nowcast error at $t+h$, and $\mu_{t+h|t+i}$ measures the update in the forecast of y_{t+h} at time $t+i$.[9]

Our goal is to construct term structures of expectations and uncertainty for horizons $h = 0, 1, \ldots H$. We will also refer to the H-step-ahead forecast, $y_{t+H|t}$, as the long-run forecast.[10] In an extension of the CMM framework, we also track the change in long-run forecasts from one quarter

[9] 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 (under quadratic loss and stationarity). Coibion and Gorodnichenko (2015) map out the implications of different theories of imperfect information for serial correlation in forecast updates.

[10] In case of our application to the US SPF, the longest horizon H equals 15 quarters.

to the next, which is denoted as μ_t^*:

$$\mu_t^* \equiv y_{t+H|t} - y_{t+H-1|t-1}. \tag{4}$$

Without additional assumptions about the dynamic processes for e_{t+h}, $\mu_{t+h|t+i}$, or μ_t^*, equations (1) and (4) represent mere accounting identities (with equation (1) also used by CMM). We collect $\mu_{t+h|t}$ (for all $0 \leq h \leq H$), μ_t^*, and e_{t-1} in a vector denoted $\boldsymbol{\eta}_t$.[11] For brevity, we refer to $\boldsymbol{\eta}_t$ as the vector of forecast updates.

Applied to forecasts $y_{t+h|t}$ for all $h \geq 0$, the accounting identities above lead to the following recursive representation of the state vector $\boldsymbol{Y}_t$:

$$\boldsymbol{Y}_t = \boldsymbol{F}\,\boldsymbol{Y}_{t-1} + \boldsymbol{\eta}_t, \tag{5}$$

where

$$\underbrace{\begin{bmatrix} y_{t-1} \\ y_{t|t} \\ y_{t+1|t} \\ \vdots \\ y_{t+H|t} \end{bmatrix}}_{\boldsymbol{Y}_t} = \underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & \dots & & 0 & \ddots & 0 \\ 0 & \dots & & \dots & 0 & 1 \\ 0 & \dots & & \dots & 0 & 1 \end{bmatrix}}_{\boldsymbol{F}} \boldsymbol{Y}_{t-1} + \underbrace{\begin{bmatrix} e_{t-1} \\ \mu_{t|t} \\ \mu_{t+1|t} \\ \vdots \\ \mu_t^* \end{bmatrix}}_{\boldsymbol{\eta}_t}.$$

The state vector $\boldsymbol{Y}_t$ contains the lagged value of y, whose first reading typically becomes available only in the current quarter, and quarterly SPF point forecasts from period t through $t+H$ (some or many unobserved). Equation (5) describes the evolution of forecasts and realized values, for a given sequence of forecast updates, $\boldsymbol{\eta}_t$. Critically, the transition matrix $\boldsymbol{F}$ is known and need not be estimated.[12] To close our model, we need to specify the dynamics of the forecast update vector $\boldsymbol{\eta}_t$, which we consider in the next subsection.

[11] The vector $\boldsymbol{\eta}_t$ includes the lagged nowcast error, since the realized value y_{t-1} is observed only at time t. CMM defined the same vector of forecast updates, except for our inclusion of the newly defined change in the long-run forecast, μ_t^*, and longer forecast horizons.

[12] All eigenvalues of the transition matrix $\boldsymbol{F}$ are 0, except for a single unit root.

Although our model is written with a state vector $\boldsymbol{Y}_t$ containing SPF forecasts, we should emphasize that, in our baseline model, 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. Assuming $\boldsymbol{\eta}_t$ is (mean-) stationary, our state equation embeds a common trend assumption, whereby survey forecasts and outcome variables share a single common trend, and survey forecast errors are (mean-) stationary. In addition, variations in the common trend component depend on the size of variations in μ_t^*. When μ_t^* is 0 our model includes the case in which all elements of $\boldsymbol{Y}_t$ are stationary (and for arbitrarily small variations in μ_t^*, the model provides a near-stationary representation of the data).

4.2 Transition equation in the martingale case

In our baseline specification, we embed an insight also underlying the original model of CMM: If point forecasts are optimal (under quadratic loss), then forecasts are martingales. As a consequence, forecast updates or revisions from one forecast origin to the next will be martingale differences. We thus assume that the expectational update $\mu_{t+h|t}$, forms a martingale difference sequence (MDS): $E_{t-1}\mu_{t+h|t} = 0$, and similarly for the nowcast error $E_{t-1}e_t = 0$.[13] In addition, we assume that changes in the long-run forecast are martingale differences, $E_{t-1}\mu_t^* = 0$. The latter assumption treats the long-run forecast, $y_{t+H|t}$, as the Beveridge-Nelson trend of y_t.[14] All told, our baseline model treats the vector $\boldsymbol{\eta}_t$ as a martingale difference sequence, $E_{t-1}\boldsymbol{\eta}_t = \mathbf{0}$. Throughout, we also assume that $\boldsymbol{\eta}_t$ is Gaussian (at least conditionally on yet-to-be defined time-varying parameters).

In our baseline specification the state-space model has the following form:

$$\boldsymbol{Y}_t = \boldsymbol{F}\,\boldsymbol{Y}_{t-1} + \boldsymbol{\eta}_t\,, \tag{5}$$

$$\boldsymbol{Z}_t = \boldsymbol{C}_t\,\boldsymbol{Y}_t\,, \tag{6}$$

$$\boldsymbol{\eta}_t \sim N\left(\mathbf{0}, \mathrm{Var}_t\,\boldsymbol{\eta}_t\right)\,, \tag{7}$$

[13] We use the expectations operator E_t to denote true expectations, under the average SPF respondent's information set, which is assumed to contain the econometrician's information set at time t. In the MDS case, observed SPF forecasts are assumed to be identical to true expectations, $y_{t+h|t} = E_t y_{t+h}$.

[14] As noted before, the relative variance of μ_t^* compared to the other innovations contained in $\boldsymbol{\eta}_t$ determines the strength of any non-stationary component in the data.

with the evolution of $\text{Var}_t \boldsymbol{\eta}_t$ as a stochastic volatility (SV) process to be defined further below.

4.3 Measurement equation

To explain the measurement equation, we need additional notation for forecasts, covering the different types of annual forecasts (simple annual averages for the unemployment rate and percent changes in annual averages for GDP growth and inflation in the GDP price index). To match forecasts of annual average levels and their growth rates, let

$$\bar{y}_t = 1/4 \cdot \sum_{j=0}^{3} y_{t-j}, \tag{8}$$

$$\hat{y}_t = 1/16 \cdot (y_t + 2 \cdot y_{t-1} + 3 \cdot y_{t-2} + 4 \cdot y_{t-3} + 3 \cdot y_{t-4} + 2 \cdot y_{t-3} + y_{t-2}) \ , \tag{9}$$

so that when t corresponds to a Q4 observation, $\bar{y}_t$ and $\hat{y}_t$ denote an annual observation.[15] The formula for $\hat{y}_t$ represents a log-linear approximation to the growth rate in average levels of the years (or four-quarter intervals) ending at t and $t-4$. Other examples relying on such approximations include Aruoba (2020), Mariano and Murasawa (2003), and Patton and Timmermann (2012).

The variables $y_{t+h|t}$, $\bar{y}_{t+h|t}$, and $\hat{y}_{t+h|t}$ denote survey expectations collected at forecast origin t for forecast targets y_{t+h}, $\bar{y}_{t+h}$, and $\hat{y}_{t+h}$, respectively. At a given point in time, t, the survey data is assumed to provide observations of $y_{t+h|t}$, $\bar{y}_{t+h|t}$, and/or $\hat{y}_{t+h|t}$, for different (but separate) values of $h \geq 0$.[16] At different points in time, survey forecasts for different horizons h may be available. As noted above, the longest forecast horizon H covered by our model reflects the availability of annual forecasts. The quarterly fixed-horizon forecasts from the SPF extend only 4 quarters beyond the origin t (so this is the maximum horizon used by CMM). By making use of annual forecasts for up to three calendar years out, our term structures cover 16 quarters, ranging from the current quarter, $h = 0$, to $H = 15$.

The measurement vector $\boldsymbol{Z}_t$ of the model contains readings from the SPF about forecasts for fixed horizons, $y_{t+h|t}$, or fixed events, $\bar{y}_{t+h|t}$ and $\hat{y}_{t+h|t}$, respectively, as well as a reading of the last realized value, y_{t-1}, that are available at a given point in time t for different horizons h. Since the

[15] The use of a scale factor of 16 in the denominator of equation (9) reflects our definition of y_t as an *annualized* rate of change at quarterly frequency. Analogously to derivations known from Mariano and Murasawa (2003), if y_t was defined to measure a quarterly rate of change, the appropriate scale factor would be 4.

[16] Typically, a given survey source, like the (US) SPF, provides readings on $\bar{y}_{t+h|t}$ or $\hat{y}_{t+h|t}$, plus (possibly) $y_{t+h|t}$.

SPF provides fixed-horizon forecasts for up to four quarters ahead, we disregard all current-year (fixed-event) forecasts. In addition, when the SPF is conducted in the fourth quarter of the current year, we also disregard the next-year forecast.[17] Otherwise, we include in $\boldsymbol{Z}_t$ all available readings of fixed-event forecasts for the next year and beyond. With this specification, the measurement equation takes the form

$$\boldsymbol{Z}_t = \boldsymbol{C}_t \boldsymbol{Y}_t,$$

where the elements of $\boldsymbol{C}_t$ are time-varying to reflect shifting data availability across surveys and over time (with the SPF adding additional years of annual forecasts as indicated above) but known.

Our approach differs in some notable ways from some related work. Assuming rational expectations, one approach for fitting an arbitrary term structure of survey forecasts is to assume a time-series model for the evolution of y_t that generates forecasts $E_t y_{t+h}$ and to assume that observed survey forecasts are equal to model-implied forecasts plus a measurement error,

$$y_{t+h|t} = E_t y_{t+h} + \text{noise}_t \,,$$

and likewise for $\bar{y}_{t+h|t}$ and $\hat{y}_{t+h|t}$. This approach is employed by studies such as Aruoba (2020), Crump, et al. (2021), Grishchenko, Mouabbi, and Renne (2019), Kozicki and Tinsley (2012), and Mertens and Nason (2020). Potential drawbacks of this approach 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 ($E_t y_{t+h}$).[18] 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 invariably comes at the cost of discarding part of the survey respondents' judgement and broader modeling. Instead, as noted above, in our baseline model we aim to perfectly fit model-implied forecasts to survey forecasts,

[17] We compared the SPF's current year forecasts with their implied counterparts constructed from lagged data and the SPF's fixed-horizons forecasts and found any differences to be typically small, and similarly so for next-year forecasts made in the fourth quarter of a given year. While the sources of such differences may otherwise be of independent interest, they do not appear germane to our study.

[18] Survey respondents (or the average survey respondent) may generate persistent forecast errors, consistent with information frictions. For simplicity, this case is ignored in our baseline specification. However, as detailed in Section 4.5, our approach can readily be extended to handle the case of persistent survey errors using a VAR specification of the $\boldsymbol{\eta}_t$ vector as in Clark, McCracken, and Mertens (2020).

$y_{t+h|t} = E_t y_{t+h}$, and our extended model sees differences between forecasts from SPF and model as bias (and thus predictable forecast errors, instead of noise in measurement).

4.4 Stochastic volatility

In the stochastic volatility process of the innovation vector $\boldsymbol{\eta}_t$, we distinguish the last component of $\boldsymbol{\eta}_t$ that is the change in the long-run forecasts from one quarter to the next from the other components comprised of forecast updates at shorter horizons. We assume that the change in the long-run forecasts affects all forecasts at shorter horizons, and that it has a variance that is constant over time. For all forecast horizons except the long-run one, we assume a scalar volatility process that is common to all horizons. With these assumptions, the model decomposes forecast updates into long-run shifts and cyclical gaps, as follows:

$$\begin{aligned}
\boldsymbol{\eta}_t &= \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} + \mathbf{1} \cdot \boldsymbol{\mu}_{\boldsymbol{t}}^{*}, \qquad (10)\\
\boldsymbol{\mu}_{\boldsymbol{t}}^{*} &\sim N(0, \sigma_{*}^{2}),\\
\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} &\sim N\left(0, \lambda_t \cdot \tilde{\boldsymbol{\Sigma}}\right),\\
\ln \lambda_t &= \delta \ln \lambda_{t-1} + \nu_t,\ \nu_t \sim N(0, \sigma_{\nu}^{2}). \qquad (11)
\end{aligned}$$

Carriero, Clark, and Marcellino (2016) develop a common volatility model for use with VARs in which volatility comovement is high.[19] In the estimates of CMM, the comovement of volatility in SPF forecast updates was high across quarterly forecast horizons extending four periods ahead. In this paper, we impose the common volatility specification in part for dimension reduction in a larger model and in part to improve estimation efficiency for longer forecast horizons for which only annual forecasts are available and even then not at all forecast origins. In addition, when the model is estimated for SPF predictions of a scalar outcome variable, y_t, commonality in uncertainty across forecast horizons appears a suitable assumption.

In our implementation, to simplify the treatment or identification of variable scales, we specify the log SV process to have a mean of 0, reflected in the AR(1) process of equation (11) that omits an intercept. The time-varying variance with mean of 1 scales up a full variance-covariance matrix

[19] Chan (2020) extends the model to include other features, including non-Gaussian and serially dependent innovations, and develops a computationally faster algorithm.

$\tilde{\boldsymbol{\Sigma}}$. Although the ordering of variables commonly affects estimates of VARs with SV processes for each variable (see discussions in studies such as Arias, Rubio-Ramirez, and Shin (2022)), with our common SV specification, the ordering of variables in the model has no impact on estimates.

4.5 Non-MDS generalization

To allow for possible biases and persistence in forecast errors and expectational updates, we also consider an extension of our model that does not rest on the MDS assumption. In this extended model, we allow for differences to emerge between model-implied forecasts, $E_t y_{t+h}$, and the term structure of SPF-consistent expectations, $y_{t+h|t}$, that is tracked by the state vector $\boldsymbol{Y}_t$.[20] Specifically, in this model, we continue to treat the long-run forecast as a random walk such that μ_t^* is mean 0, but we let $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ follow a VAR(1):[21]

$$\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} = \tilde{\boldsymbol{G}}\tilde{\boldsymbol{\eta}}_{\boldsymbol{t-1}} + \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}}. \tag{12}$$

In this case, the state equation becomes:

$$\boldsymbol{Y}_t = (\boldsymbol{G} + \boldsymbol{F})\,\boldsymbol{Y}_{t-1} - \boldsymbol{G}\boldsymbol{F}\,\boldsymbol{Y}_{t-2} + \boldsymbol{\varepsilon}_t, \tag{13}$$

$$\text{where} \quad \boldsymbol{G} = \boldsymbol{H}\tilde{\boldsymbol{G}}\boldsymbol{H}^{-1}, \qquad \text{with} \quad \boldsymbol{H} = \begin{bmatrix} \boldsymbol{I} & \mathbf{1} \\ \mathbf{0} & 1 \end{bmatrix}, \qquad \text{and} \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \\ \mu_t^* \end{bmatrix}. \tag{14}$$

For the VAR shocks, we assume an SV specifications that is analogous to what was used for $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ in the non-MDS case, $\tilde{\boldsymbol{\varepsilon}}_{\boldsymbol{t}} \sim N\left(\mathbf{0}, \lambda_t \cdot \tilde{\boldsymbol{\Sigma}}\right)$, and we also continue to assume $\boldsymbol{\mu}_{\boldsymbol{t}}^{*} \sim N(0, \sigma_*^2)$.[22]

In the non-MDS case, we still rely on the recursive representation of $\boldsymbol{Y}_t$ as driven by a sequence of forecast updates in equation (5). Without the strict MDS assumption, equation (5) can be viewed as a prewhitening step that at least reduces (but does not fully eliminate) persistence in the data. Moreover, when $\tilde{\boldsymbol{G}} = \mathbf{0}$, the model nests the MDS case. By centering a Bayesian prior

[20] We refer to the set of predictions, $y_{t+h|t}$, contained in $\boldsymbol{Y}_t$ as the term structure of "SPF-consistent" expectations, since $\boldsymbol{Y}_t$ has to be (at least partially) interpolated from observed SPF forecasts for various fixed horizons and events.

[21] Lag order larger than one could also be considered. However, CMM already found little support for such choices.

[22] The representation of $\boldsymbol{Y}_t$ as VAR(2) in equation (13), rather than in companion form, is useful for efficient implementation with a precision-based sampler.

around $\tilde{\boldsymbol{G}} = \mathbf{0}$, estimation of the non-MDS model can embody prior beliefs over the (squared-loss) efficiency of the SPF. Indeed, in our Bayesian estimation, we place such a prior (with Minnesota-style shrinkage) on $\tilde{\boldsymbol{G}}$.

Critically, forecasts of y_t derived from the non-MDS model generally differ from the SPF, as the non-MDS model allows for serially correlated forecast updates, and thus predictable SPF forecast errors. Differences between model estimates of $E_t y_{t+h}$ and the term structure of SPF-consistent forecasts, measured by $y_{t+h|t}$, reflect the model's assessment of bias in the SPF.

4.6 Homoskedastic specifications

While our baseline model and its non-MDS generalization feature stochastic volatility, we include in our empirical analysis comparisons to estimates from models simplified to treat the innovation vector $\boldsymbol{\eta}_t$ as conditionally homoskedastic. In one specification that embeds the MDS assumption, referred to as CONST below, $\boldsymbol{\eta}_t \sim N(0, \boldsymbol{\Sigma})$. In the non-MDS generalization including conditional homoskedasticity, we allow $\boldsymbol{\eta}_t$ to follow a VAR(1) with an innovation vector $\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma})$ and transition matrix $\boldsymbol{G}$.[23] Otherwise, the state equation remains as in equation (13).

4.7 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.[24]

In the baseline specification, the objects to be estimated include the δ and σ_ν^2 parameters of the SV process, the constant innovation covariance matrix $\tilde{\boldsymbol{\Sigma}}$, the variance of innovations to the long-run forecast σ_*^2, the time series of the latent volatility state λ_t, and the time series of latent forecast states contained in $\boldsymbol{Y}_t$. We sample the parameters of the SV process using a conventional Gaussian prior and conditional posterior for δ and a standard inverse Gamma prior and conditional posterior for σ_ν^2. We sample $\tilde{\boldsymbol{\Sigma}}$ with an inverse Wishart prior and conditional posterior. We draw

[23] Compared to the SV case, the CONST specification of the non-MDS model also allows for persistent changes in the long-run forecast, since μ_t^* is included in the VAR.

[24] CMM also considered joint estimation of their model for several economic variables, but without consistent gains in forecast performance or other notable changes in estimates.

σ_*^2 with an inverse Gamma prior and conditional posterior. We estimate the volatility state λ_t with the mixture approach of Kim, Shephard, and Chib (1998), as refined in Omori, et al. (2007).

Sampling the latent forecast states contained in $\boldsymbol{Y}_t$ involves more complexity, and we use a new precision-based sampler to efficiently tackle these complexities; for details see Mertens (2022). In conventional problems, the sampling of the latent state vector can be handled with a simulation smoother such as that of Durbin and Koopman (2002). However, in our setting, in which our model assumes no measurement error, we face an ill-defined posterior precision: $|\operatorname{Var}(\boldsymbol{Y}_t|\boldsymbol{Z}^t)| = 0$. To efficiently draw $\boldsymbol{Y}^t|\boldsymbol{Z}^t$ in this case, we use a new precision-based sampler. Originally developed in Chib and Jeliazkov (2006) and then Chan and Jeliazkov (2009), precision-based samplers have been used in VAR settings in studies such as Chan (2020). 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.

4.8 Entropic tilting

Introduced to the forecasting literature by Robertson, Tallman, and Whiteman (2005), entropic tilting has gained popularity as a convenient post-estimation method to incorporate additional moment conditions into a model's predictive distribution; see, e.g., Cogley, Morozov, and Sargent (2005), Ganics and Odendahl (2021), and Krüger, Clark, and Ravazzolo (2017). In our application, the moment restrictions come from the fixed-event SPF probability bin forecasts, which we recognize as potentially useful information that was not utilized to estimate the model described in Sections 4.1 through 4.4. At the same time, we do not want to deviate "too much" from the model's predictions. Entropic tilting achieves these objectives simultaneously by re-weighting the draws from the model's predictive distribution such that the re-weighted draws satisfy the moment restrictions and the new distribution is closest to the original one in the Kullback–Leibler sense.

Let $\boldsymbol{X}_t = \left[x_t^1, \ldots, x_t^n, \ldots, x_t^N\right]'$ collect predictions at forecast origin t for *calendar* years $1, \ldots, N$. For example, superscript 1 corresponds to the next calendar year. Depending on the underlying variable, x_t^n refers to the calendar-year *average level* (in the case of the unemployment rate) or the calendar-year *rate of change in yearly average levels* (in the case of GDP growth and inflation in the GDP price index).

To construct predictive densities for calendar year events, we need to distinguish between the case of the unemployment rate and the case of growth rates for real GDP and its price index. In the case of the unemployment rate, the SPF histograms solicit predictions for annual average levels, so that equation (8) applies in a straightforward manner. In the case of GDP and its price index, the SPF histograms reflect predictions for rates of change in annual average levels. To map draws from the predictive density of our quarterly model into annual growth rates, let $\mathcal{Y}_t$ denote the level of GDP (or its price index) in quarter t.[25] As before, $y_{t+h} = 400(\log \mathcal{Y}_{t+h} - \log \mathcal{Y}_{t+h-1})$ denotes the annualized quarterly growth rate of GDP from quarter $t-1$ to t and we let $R_{t+h} \equiv \exp(y_{t+h}/400) = \mathcal{Y}_{t+h}/\mathcal{Y}_{t+h-1}$ denote the corresponding quarterly gross rate of change. For simplicity, consider calculating the growth rate of average annual GDP (or its price index) from the current *calendar* year to the next *calendar* year. Standing at forecast origin t, let τ denote the last quarter of the previous calendar year, so that $\tau+1, \tau+2, \tau+3, \tau+4$ point to the four quarters of the current year, and $\tau+5, \tau+6, \tau+7, \tau+8$ refer to the next calendar year's quarters. With this notation, the SPF's concept of next year's (fixed-event) annual growth can be expressed as

$$x_t^1 = 100 \cdot \left(\frac{\mathcal{Y}_{\tau+5} + \mathcal{Y}_{\tau+6} + \mathcal{Y}_{\tau+7} + \mathcal{Y}_{\tau+8}}{\mathcal{Y}_{\tau+1} + \mathcal{Y}_{\tau+2} + \mathcal{Y}_{\tau+3} + \mathcal{Y}_{\tau+4}} - 1 \right), \tag{15}$$

$$= 100 \cdot \left(\frac{\mathcal{Y}_{\tau+5} \left(1 + \frac{\mathcal{Y}_{\tau+6}}{\mathcal{Y}_{\tau+5}} + \frac{\mathcal{Y}_{\tau+7}}{\mathcal{Y}_{\tau+5}} + \frac{\mathcal{Y}_{\tau+8}}{\mathcal{Y}_{\tau+5}}\right)}{\mathcal{Y}_{\tau+1} \left(1 + \frac{\mathcal{Y}_{\tau+2}}{\mathcal{Y}_{\tau+1}} + \frac{\mathcal{Y}_{\tau+3}}{\mathcal{Y}_{\tau+1}} + \frac{\mathcal{Y}_{\tau+4}}{\mathcal{Y}_{\tau+1}}\right)} - 1 \right), \tag{16}$$

$$= 100 \cdot \left(\left(\prod_{j=2}^{5} R_{\tau+j} \right) \cdot \frac{1 + \sum_{k=2}^{4} \prod_{j=1}^{k} R_{\tau+4+j}}{1 + \sum_{k=2}^{4} \prod_{j=1}^{k} R_{\tau+j}} - 1 \right). \tag{17}$$

Based on equation (17), we map the predictive distribution of x_t^1 into the model as follows: For $\tau + j < t$, we take observed vintage data for $R_{\tau+j}$ that was available to the SPF forecaster at t, and for $\tau + j \geq t$ we generate draws of $R_{\tau+j}$ from the model. For the remaining elements of $\boldsymbol{X}_t$, analogous computations are applied to predictions for two- and three-calendar-years ahead. Similar to the calculations of Clements (2018), our mapping from model-implied densities to the SPF histograms is thus free of log-linear approximations for annual growth rates.[26]

[25] For brevity, our description will henceforth refer to the case of real GDP, with analogous calculations applicable in the case of the GDP price index.)

[26] As described in Section 4.3, prior to the application of entropic tilting, MCMC estimation of the model for GDP and its price index invariably relies on a log-linear approximation for mapping observed SPF point forecasts into the state space.

Let f_t denote the predictive distribution of $\boldsymbol{X}_t$, given in the form of draws $\{\boldsymbol{X}_t^s\}_{s=1}^S$ obtained from MCMC estimation of the model, each with corresponding probability $w_s = 1/S$. In addition, let $\{\widetilde{w}_s\}_{s=1}^S$ denote an alternative set of weights (for the same draws) that characterizes the distribution $\widetilde{f}_t$. The Kullback–Leibler divergence (KL) between f_t and $\widetilde{f}_t$ is then defined as

$$\mathrm{KL}\left(\widetilde{f}_t, f_t\right) \equiv \sum_{s=1}^{S} \widetilde{w}_s \log\left(\frac{\widetilde{w}_s}{w_s}\right) . \tag{18}$$

Our use of entropic tilting seeks a distribution $\widetilde{f}_t$ that matches the information obtained from the fixed-event SPF probability bins while staying as close as possible, in terms of KL, to the original, model-based distribution f_t. Formally, entropic tilting is a constrained minimization with KL as the objective function and subject to constraints that represent the SPF bins as moment conditions (which f_t does not satisfy in general). Before providing a formal definition of the entropic tilting problem, we discuss the mapping between the discrete bins of the SPF histograms and our model output.

The SPF probability forecasts are available as probabilities assigned to pre-specified bins. For example, in the 2021Q4 SPF round, panelists were asked to provide the probability of the unemployment rate falling into the following 10 bins: < 3, 3 to 3.9, 4 to 4.9, 5 to 5.9, 6 to 6.9, 7 to 7.9, 8 to 9.9, 10 to 11.9, 12 to 14.9, and ≥ 15, all in average annual percentage points between calendar years 2021 and 2024. Two remarks are in order. First, the SPF bins are not literally contiguous, since to the right of each inner bin, there is a 10-basis-point-wide gap (consistent with the use of data rounded to the first decimal — however, we do not use such rounding of the data in our analysis). When mapping model-implied, continuous densities to the bins, we assign half of these gaps to bins on either side. For example, we interpret the second and third bins as $b_2 = [3.0, 3.95)$ and $b_3 = [3.95, 4.95)$, respectively. Second, while SPF panelists also submit predictions for the current calendar year, we are not utilizing this information, since for most quarters, good parts of the year are already locked in as data, leading to particularly narrow densities for current *calendar-year* predictions. Let us collect the probabilities we use in a vector $\bar{\boldsymbol{p}}_t$. The dimension of $\bar{\boldsymbol{p}}_t$ is the product of (a) the number of calendar-year predictions we utilize for entropic tilting and (b) the number of bins less one (one bin can be dropped without loss of generality, as probabilities sum to one for a given calendar-year target). Staying with the example, in 2021Q4, we use calendar-year

forecasts for $N = 3$ years (namely, 2022, 2023, and 2024), and $B = 9$ bins for each, leading to $\dim(\bar{\boldsymbol{p}}_t) = N \cdot B = 27$.

Probabilities can be written as expectations over indicator functions, and thus as moments. Specifically, for event $\mathcal{A}$ we have $\mathbb{P}(\mathcal{A}) = \mathrm{E}(\mathbb{1}(\mathcal{A}))$, where $\mathbb{1}(\cdot)$ is the indicator function. Hence, by defining $p \equiv \left[p_1, \ldots, p_n, \ldots, p_N\right]'$, where $p_n(x_t^n) = \left[\mathbb{1}(x_t^n \in b_1), \ldots, \mathbb{1}(x_t^n \in b_B)\right]'$, we can impose the probabilities in the SPF bin forecasts as moments on the tilted distribution. Let us collect the functions and moment conditions as $g \equiv p$ and $\bar{\boldsymbol{g}}_t \equiv \bar{\boldsymbol{p}}_t$, respectively, where $g : \mathbb{R}^N \to \mathbb{R}^{\dim(\bar{\boldsymbol{g}}_t)}$.

Entropic tilting ensures that under the tilted distribution $\widetilde{f}_t$, we have $\mathrm{E}_{\widetilde{f}_t} g(\boldsymbol{X}_t) = \bar{\boldsymbol{g}}_t$, and $\mathrm{E}_{\widetilde{f}_t} g(\boldsymbol{X}_t) \equiv \sum_{s=1}^{S} \widetilde{w}_s g(\boldsymbol{X}_t^s)$ denotes the expected value of $g(\boldsymbol{X}_t)$ under the distribution $\widetilde{f}_t$. Entropic tilting is then the solution of the following optimization problem:

$$\min_{\widetilde{f}_t \in \mathbb{F}_t} \mathrm{KL}(\widetilde{f}_t, f_t) \qquad \text{such that} \qquad \mathrm{E}_{\widetilde{f}_t} g(\boldsymbol{X}_t) = \bar{\boldsymbol{g}}_t \,, \tag{19}$$

where $\mathbb{F}_t$ is the set of feasible distributions given the predictive draws for f_t.[27] The solution to this problem is given by

$$\widetilde{w}_s^* = \frac{\exp\left(\gamma^{*\prime} g(\boldsymbol{X}_t^s)\right)}{\sum_{s=1}^{S} \exp\left(\gamma^{*\prime} g(\boldsymbol{X}_t^s)\right)} \,, \tag{20}$$

with

$$\gamma^* = \mathrm{argmin}_\gamma \sum_{s=1}^{S} \exp\left(\gamma' \left(g(\boldsymbol{X}_t^s) - \bar{\boldsymbol{g}}_t\right)\right) \,, \tag{21}$$

where $\{\widetilde{w}_s^*\}_{s=1}^S$ denotes the new weights used to construct the tilted predictive distribution.

Our use of entropic tilting means that we do not need to either fit parametric distributions to the SPF histograms or make semi-parametric assumptions to make use of them, unlike much of the literature referenced by Bassetti, Casarin, and Del Negro (2022). We see this as an advantage of our approach, adhering to the information contained in the SPF. Furthermore, to the best of our knowledge, we are the first to incorporate histogram bins directly into entropic tilting, except for mention of the approach in the documentation of the Bayesian Estimation, Analysis, and Regression toolbox (BEAR) maintained at the ECB (Dieppe, Legrand, and van Roye, 2016).

[27] As noted earlier, the model tends to deliver fairly tight distributions for the current year's forecasts, while the SPF probability bin predictions often put positive mass on bins where there is no draw from the model's predictive distribution. In such cases, the SPF's distribution is not in the set of feasible distributions $\mathbb{F}_t$.

5 Results

This section first documents the historical time variation in the volatility of SPF forecast errors. We then present estimates of the term structures of expectations using our baseline model, followed by fan charts for selected forecast origins. The section then turns to the impacts of entropic tilting on forecasts, forecast uncertainty, and historical forecast accuracy. It next compares forecast accuracy for our baseline model that exploits MDS assumptions to the model that does not do so. The section then provides selected results for the unemployment rate and inflation. Finally, the section presents a practical application using our estimated term structure of forecasts: constructing fan charts with calendar year (four-quarter growth rates) forecasts, in the style used by the FOMC.

5.1 Time variation in volatility

To show a key factor in our rationale for including stochastic volatility in the model, Figure 1 provides time-varying quarterly volatility estimates obtained with the expectational updates for GDP growth for selected horizons, including the short horizons of $h = 0, 1$ for which the expectational updates are directly observable and longer horizons of $h = 7, 11$ (corresponding to two and three years ahead) for which quarterly expectational updates are not directly observable and are instead inferred from the model and observed annual forecasts.[28] Specifically, the black lines provide the full-sample (smoothed) estimates of time-varying volatility.[29] While not shown in the interest of chart readability, real-time estimates of stochastic volatility (obtained by looping over time and estimating a historical volatility path at each forecast origin) that underlay the forecast results considered in the next section are very similar to the full-sample estimates. For comparison to the volatility estimates, the figures include (in light-shaded bars) the absolute values of the expectational updates, which roughly correspond to the objects that drive the model's volatility estimates.

Consistent with the findings of CMM, the volatility estimates display several broad features (although we only show estimates for GDP growth, estimates for other variables share the same features). First, the time variation in volatility is considerable. In the period preceding the tem-

[28] The directly observable elements of $\boldsymbol{\eta}_t$ for $h = 0, \ldots, 3$ were also used as data in the CMM model.

[29] In terms of the common-volatility specification described in Section 4.4, time-varying volatilities associated with different elements of $\boldsymbol{\eta}_t$ are given by the square roots of the corresponding diagonal elements of $\lambda_t \cdot \tilde{\boldsymbol{\Sigma}}$ as defined in equation (10).

porary surge of volatility induced by the COVID-19 outbreak, the highs in the volatility estimates are typically 3 to 4 times the levels of the lows in the estimates. As documented in other studies, such as Carriero, et al. (2021), volatility soared to unprecedented levels in the first few quarters of the pandemic. Second, some of the time variation occurs at low frequencies, chiefly with the Great Moderation of the 1980s. But some of the time variation is cyclical, as volatility has some tendency to rise temporarily around recessions. For example, the volatility of GDP growth rises with most recessions. Third, by construction, the volatility estimates obtained from our common-SV specification display perfect comovement across horizons. However, the observable data for elements of $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ share this commonality.

[Figure 1 about here.]

5.2 Term structures of expectations

To illustrate the term structures of expectations produced by our model, Figure 2 presents estimates of point forecasts for GDP growth, using SPF forecasts available as of the indicated forecast origins. As detailed above, the model uses as inputs quarterly SPF forecasts up through the 4-steps-ahead horizon and annual fixed-event forecasts that vary in their horizon availability (or what we use with Q4 forecast origins); the model yields estimated quarterly forecasts through the 15-steps-ahead horizon. In each panel, red diamonds give the available quarterly fixed-horizon forecasts from SPF, and dotted lines give the available annual fixed-event forecasts that bear on the indicated quarter, with current year in blue, year ahead in green, and two years ahead in red. Because the annual forecasts are percent changes in annual averages of GDP, the annual forecasts reflect information in quarterly growth rates for 7 quarters, as indicated in our previous discussion of the model's measurement equations, in equation (9). The black lines provide the real-time quarterly SPF-consistent forecast estimates, with 68 percent credible sets indicated with the gray dotted lines.

One pattern in the results reflects the design of the model: The model yields quarterly forecasts for horizons 0 through 4 quarters that match exactly the reported SPF projections, with no uncertainty around the estimate. Only the model-based forecasts for subsequent quarters show any uncertainty around them.

The second notable pattern in Figure 2's results is that the estimated quarterly forecasts for horizons of 5 through 15 quarters ahead interpolate through the annual fixed-event forecasts, with some fluctuations around the annual forecasts. For horizons of 5 through 15 quarters, the black lines representing quarterly forecasts generally intersect the dots (representing the annual forecasts) within the first 3 or 4 quarters covered. For example, in 2009Q1, when the SPF reported a current-year annual forecast but not subsequent year forecasts, the model-based quarterly forecast intersects the annual forecast before the mid-point of the blue dots, and then remains above the 2009 annual year forecast for subsequent quarters. Similarly, in 2019Q4 (note that this panel shows dots for two rather than three years of annual forecasts, because due to strong overlap with quarterly forecasts, the annual forecast from the SPF for the calendar year of 2020 is not used in model estimation), the model-based quarterly forecast intersects the annual forecast for 2022 before the mid-point of the red dots, and then remains above the 2022 annual year forecast for a few quarters before moving to a lower level. In a number of the other cases shown, the model's interpolated quarterly forecasts intersect the annual forecast a second time late in the period covered by the annual prediction. This pattern applies to two of the three years covered in the 2009Q2 results, all three years in the 2017Q3 estimates, and one of the two years in the 2019Q4 results. Overall, while the model's estimates of GDP growth forecasts for quarters 5 through 15 interpolate through the available annual forecasts, they display some fluctuations.

While not reported in the paper in the interest of brevity, estimates for the unemployment rate provided in the supplementary online appendix show two other patterns that either differ modestly from or are not evident in the GDP results. First, for this variable the annual forecasts are simply annual averages of forecasts within the four quarters of the year, and the model's interpolation of quarterly forecasts tend to show less variation. For horizons of 5 quarters and beyond, the estimated quarterly forecasts typically intersect the annual forecasts once, not twice. Second, with the availability of annual forecasts at longer horizons more mixed for the unemployment rate, the estimated term structures of forecasts for the unemployment rate indicate that the availability of annual forecasts at longer horizons tends to increase the precision of the quarterly forecast estimates. That is, the quarterly estimates tend to be more precise when annual forecasts for the quarters are available than when they are not available.

[Figure 2 about here.]

5.3 Fan charts and term structures of uncertainty

After illustrating the basic properties of our model regarding the point forecasts, we now add forecast uncertainty to the picture. For this purpose, we provide results in fan charts for GDP growth, in Figures 3 and 4. The fan charts provide SPF-consistent quarterly forecasts for horizons through 15 quarters ahead, including the point forecast and the predictive density as captured by 68 percent bands. The entirely model-based results are reported in red; the black line and shaded region provide results that include entropic tilting to the SPF's probability bins, which are discussed in Section 5.4 below. These results all correspond to real-time forecasts, with actual outcomes given in green.

Figure 3 provides results for growth using the second and fourth quarters from a few years selected from a period spanning the end of the Great Recession through the ensuing gradual recovery. These results show two general patterns. First, the width of the uncertainty bands appears to vary over time. In particular, uncertainty appears to be greater in 2013Q2 than 2015Q4. Second, as expected, forecast uncertainty tends to rise with the forecast horizon. In the case of GDP growth, the widths of the uncertainty bands increase significantly from the current quarter through about the 8-steps-ahead horizon and then stabilize somewhat (but continue to drift up, as will be shown below). While not shown in the interest of brevity, for a more persistent variable such as the unemployment rate, the widths of the uncertainty bands in similar fan charts increase steadily across all horizons.

[Figure 3 about here.]

Figure 4 provides results for GDP growth using a few quarters from the period of the COVID-19 pandemic. When the 2020Q1 SPF was collected and published in early to mid-February, the virus had emerged but not drawn much attention or had any economic impact in the US. Of course, that quickly changed. Actual GDP growth plummeted from 2020Q1 to 2020Q2. As shown in Figure 4, our model's estimate of uncertainty for GDP growth sharply increased from the 2020Q1 to 2020Q2 SPF, indicated by the dramatically wider density bands of the latter (top right panel) as compared to the former (top left panel). After the recovery had set in, by 2020Q4 (bottom left panel), the

model's estimates of uncertainty for GDP growth narrowed sharply. But even a year later, in 2021Q4 (bottom right panel), forecast uncertainty as captured by the width of the density's bands remained considerably greater than in the last pre-pandemic forecast, from 2020Q1.

[Figure 4 about here.]

To more directly assess changes in uncertainty across horizons and over time, Figure 5 depicts the term structure of uncertainty around quarterly GDP growth forecasts, from 1992 to 2021. For constructing the figure, uncertainty is measured by the width of the 68 percent bands of the model's predictive densities estimated in real time. For readability, the chart includes a subset of quarterly horizons, including a few short-horizon horizons and longer horizons at increments of 3 or 4 quarters. The overall increases and decreases shown in these figures mirror our earlier description of fan charts at selected forecast origins. Uncertainty is noticeably higher at longer horizons (8 or more quarters) than shorter horizons (0 to 4 quarters). From 0 to 4 quarters, uncertainty gradually increases. From 8 to 15 quarters, uncertainty continues to rise, but typically by less than in the short horizon case. Consistent with the patterns in full sample SV estimates discussed above, the uncertainty of out-of-sample forecasts of GDP growth fluctuates significantly over time. In the period since 1992, it rose some following the 2001 recession and more notably around the Great Recession and again a few years into the ensuing recovery. Then the outbreak of COVID-19 produced an unprecedented, but temporary, spike in uncertainty in 2020. In results presented in the supplementary online appendix in the interest of brevity, estimates of uncertainty that incorporate entropic tilting to the SPF probability bins are similar to the purely model-based estimates reported here, except that in the early period of the pandemic, the information from the bins helps mitigate the rise in uncertainty that occurs as macroeconomic volatility temporarily soared.

[Figure 5 about here.]

5.4 Using entropic tilting to incorporate SPF probability bins

Incorporating information from the SPF's annual fixed-event probability bins will affect the predictive densities from our model to the extent that the SPF bins differ from the purely model-based

probabilities. In our estimates, at some forecast origins, the purely model-based probabilities are comparable to the SPF bins. But to limit the volume of results, in this section we first focus on some examples in which differences are more notable, to illustrate how tilting to the SPF bins can have impacts on the model-based predictive densities. We then examine tilting's impacts on the historical accuracy of SPF forecasts.

5.4.1 Case studies of the impacts of entropic tilting to SPF bins

To gauge when tilting may matter, the SPF bins and the purely model-based forecasts can be compared in histograms using the bin definitions. The supplementary online appendix provides such an example using forecasts of the unemployment rate made in the last quarter of the Great Recession, 2009Q2. As of the 2009Q2 SPF (collected and published in May 2009), the actual unemployment rate was known through April, and given the historical persistence of unemployment, having that information made for a relatively tight predictive distribution, quite similarly for the SPF and the entirely model-based distribution (and the distributions became progressively tighter in the Q3 and Q4 surveys). As noted above, in view of the tightness of the current-year bins from SPF, we do not include the current-year histograms in our entropically-tilted forecasts. For subsequent years, the subjective SPF and model-based bins show more sizable differences. In the next year (2010) forecast, the SPF put somewhat more probability mass on high unemployment rates. Two years ahead (2011), the entirely model-based distribution was flatter than the SPF's. Three years ahead (2012), the entirely model-based distribution put significantly more weight than did the SPF on an unemployment rate below 6 percent, whereas the SPF put more weight on unemployment between 7.5 and 8.9 percent.

This section's reported results focus on GDP growth. To illustrate the impacts of entropic tilting, we rely on cumulative distribution functions (cdfs). We compute them empirically (using draws from posterior predictive distributions) for the purely model-based forecasts and for model forecasts entropically tilted to match the SPF probability bins. We report results for the baseline model with SV and its homoskedastic (CONST) counterpart; comparing across these models gives some sense of the interaction between SV (as opposed to homoskedasticity) and tilting. In the case of the SPF bins, we cumulate the histogram probabilities, treating the distribution as uniform within each interval and reporting a flat line within the range of each bin and marking the end of

the interval with a blue dot. As a last matter of explanation, the results we report are for calendar year forecasts one year ahead.

Figure 6 provides cdfs for forecasts made in 2007Q3, for the annual growth rate of GDP in 2008. As indicated in the upper left panel, the entirely model-based predictive distributions from the SV and CONST specifications display sizable differences. As compared to the cdf of CONST forecasts, the cdf of SV forecasts puts less mass in its left tail and more in the right. Both differ noticeably from the SPF bin probabilities, more sharply in the right tail for the CONST forecasts and more sharply in the left tail for the SV forecasts. Accordingly, applying entropic tilting to the model-based forecasts changes their cdfs, as shown in panels (b) and (c). In the CONST case, the tilting pulls up the model cdf, mostly in the right tail, whereas in the SV case, the tilting mostly pulls up the model cdf in the left tail. Finally, as shown in the lower right panel, the tilted cdfs from the CONST and SV models are very similar. Of course, the tilting could yield model densities that hit the moment probabilities while still distributing probability mass differently within interval of the bin. But in this example, while the tilted cdfs are not exactly the same, they are very similar, implying very similar predictive distributions for SV with entropic tilting and CONST with entropic tilting.

In another example, Figure 7 provides cdfs for forecasts made in 2013Q1, for the annual growth rate of GDP in 2014. In this case, the results in the upper left panel indicate that the entirely model-based predictive distributions from the SV and CONST specifications are quite similar. Both cdfs are aligned with the SPF bin probabilities in the left tail (albeit at the edge of the bin) but not in the right. In turn, tilting to the SPF bins rotates the model-based cdfs upward in the right tail. In this case, the tilted cdfs of the SV and CONST specifications are not just very similar but virtually indistinguishable.

[Figure 6 about here.]

[Figure 7 about here.]

The forecast fan charts for GDP growth in Figures 3 and 4 shed additional light on the effects that entropic tilting to SPF bins has on forecasts from the baseline SV specification. While the results just discussed focused on forecasts of annual growth, the fan charts illustrate the effects

that tilting (using the annual forecast probability bins) has on SPF-consistent quarterly forecasts. In Figure 3's results for selected Q2 and Q4 forecast origins in 2009, 2013, and 2015, the entropic tilting tends to pull up the point forecast from the model. For the forecast origins in which the tilting consistently lifts the point forecast, it also lifts the uncertainty bands (both lower and upper). In a number of instances, tilting to the SPF bins tends to slightly increase the width of the 68 percent uncertainty bands. But Figure 4's results for selected forecast origins in the pandemic period show that, at times, entropic tilting to the SPF probability bins narrows forecast uncertainty. As noted above, in this period, particularly in the first few quarters of the COVID-19 outbreak, extreme volatility caused stochastic volatility estimates and forecast uncertainty to rise sharply. The probability bins from the SPF (which can reflect the subjective judgment of the average survey respondent) implied less of a rise in uncertainty, such that tilting the model distributions to the bins yields less dispersion in the predictive distribution.

5.4.2 Entropic tilting's impacts on the historical accuracy of SPF forecasts

Ultimately, we are interested in whether the forecasts we construct using our model and the available quarterly and annual point forecasts from the SPF can be improved by bringing information from the SPF's annual probability bin forecasts to bear through entropic tilting. The efficacy of that additional information and tilting will depend on whether the entirely model-based predictive densities are very different from the bin forecasts and consistently better if they are different. We have provided examples in which the model-based densities differ from those implied by the SPF bins, but also noted that the differences are also sometimes small. In this section we turn to a more formal assessment of the broader question at hand. For these results, from 1992:Q1 onward (or 2009Q2 in the case of the unemployment rate, due to the availability of SPF histograms noted in Table 1), out-of-sample forecasts are generated for all quarterly horizons from $h = 0$ to 15, based on all available data since 1968Q4, by re-estimating the model at each forecast origin and simulating its predictive density. We examine both the raw forecasts from the model and those obtained by entropic tilting to the SPF's annual probability bins.

We first compare — in the most direct way possible — the accuracy of the SPF's probability bin forecasts to purely model-based forecasts, taking the histograms as they are without making the additional assumptions that would be needed to turn them into complete predictive densities. To do

so, we use the annual (calendar year) forecasts directly from the SPF, and we obtain corresponding model-based annual forecasts with the appropriate transformations of quarterly forecasts. This comparison relies on the (discrete) rank probability score (DRPS), which has been been applied in various studies of survey forecasts, including Boero, Smith, and Wallis (2011) and Clements (2018). The DRPS assigns scores based on outcomes being within bins or not, as follows:

$$\text{DRPS} = \frac{1}{P}\sum_{t=1}^{P}\sum_{k=1}^{K}\left(P_t^k - D_t^k\right)^2,$$

where P denotes the number of forecast origins (periods) included, K denotes the number of probability bins, P_t^k is the cumulative bin forecast probability from bin 1 through k, and D_t^k is the cumulation of an indicator variable with value 1 for k if the outcome falls in bin k and 0 otherwise. The lower the score, the better the forecast.

A table in the supplementary online appendix summarizes DRPS comparisons for forecasts of annual GDP growth, reporting ratios of scores for the SV and CONST models relative to scores for the SPF histogram forecasts. In results for one-year-ahead forecasts with the samples starting in 1992, the score ratios are very close to 1. Through the lens of this scoring measure, over the longer samples of available forecasts, annual probability forecasts from the model for the events covered by the SPF probability bins are no more or less accurate than the SPF's probability forecasts themselves. However, later in time, in the samples starting in 2009, the DRPS ratios exceed 1, often with statistical significance. Over these later (but also shorter) samples, the bins display more of an advantage over the purely model-based probability estimates, and more so in the sample ending before the pandemic than the one including it. In the 2009-2016 sample, it is also the case that the SPF's advantages are greater at the multi-year horizons than the one-year-ahead horizon.

To shed more light on the DRPS accuracy of the SPF bins as compared to the models, Figure 8 reports time series of GDP growth scores for the SPF and SV and CONST models (left column) and expanding window averages (right column). The last observation in the average scores corresponds to the full sample results described above (e.g., in one-year-ahead forecasts, the 1992-2020 average scores are essentially the same). These results on scores over time confirm some time variation in the relative performance of the SPF probability bins. Until about the Great Recession, the models score better than the SPF bins. But from the Great Recession until the outbreak of the pandemic,

the SPF bins were more accurate than the model-computed bins for annual forecasts. Overall, these results suggest that, relative to our models that are already centered on the SPF point forecasts, we will not find much additional payoff from the SPF annual bins (in broader forecast accuracy over long sample periods). But for forecasting GDP growth, the bins may have more useful information in the period following the Great Recession of 2007-2009 and the ensuing slow recovery. That period was unusual in the sluggishness of the recovery, which may have given an advantage to the judgment of the average SPF respondent as compared to time series models more challenged to adjust in real time to the unusual aspects of the period.

[Figure 8 about here.]

To further assess the efficacy of tilting to the SPF probability bins, we turn to quarterly forecasts of GDP growth and formally evaluate the point and density forecasts of the entropically tilted model against the predictions of the MDS model, using RMSE for point forecasts and the CRPS for density forecasts. We include results for a full sample of 1992-2022 and a sample shortened to 1992-2016 to assess possible sensitivity to the unusual outcomes from the period of the COVID-19 pandemic. Table 2 reports ratios of scores for the tilted SV forecast relative to the entirely SV model-based forecast. We also report ratios of scores for forecasts from the CONST specification and their entropically-tilted counterparts, relative to the same SV baseline. The first two columns of the table provide the raw levels of scores from the SV baseline.

Starting with point forecasts, incorporating the information in the SPF probability bins through entropic tilting has little impact on accuracy of forecasts based on the model and SPF's point forecasts. With the baseline SV model, the RMSE ratios are little different from 1 over the full sample and no lower than 0.98 over the sample ending before the pandemic. Without tilting, the point forecasts of the CONST specification are very similar in accuracy to those from the SV model (by construction, at horizons of 0 through 4 quarters, the SV and CONST forecasts exactly match the SPF forecasts). Just as tilting does not have much effect on the SV forecasts, it also does not have much effect on the CONST forecasts; the RMSE ratios (relative to baseline SV) are largely the same for CONST and CONST with tilting.

With respect to the effects of entropic tilting, the patterns in density forecast accuracy are similar to those in point forecast accuracy. In the SV results, over the full sample of 1992-2022, the

RMSE ratios for tilted versus baseline are no lower than 0.99. In the shorter sample, the CRPS ratio is 0.98 or 0.99 at a number of quarterly horizons, but none of the gains achieved by tilting are statistically significant. Larger differences in density accuracy occur in the comparison of the CONST specifications to the baseline SV model. Absent tilting, the CONST specification is less accurate than SV at many horizons, by as much as 7 percent in the pre-pandemic sample, often significantly. Applying entropic tilting to the CONST forecasts tends to very slightly improve their accuracy in the pre-pandemic sample, but not by enough to eliminate the advantage of the SV model. This finding on the benefits of SV is consistent with a number of studies noted above that have found SV to often yield improvements in forecasts from time series model.

We see the inability of entropic tilting to deliver significant gains either in point or density forecasts beyond the MDS model as an indication of the model's performance in capturing information in the SPF's point forecasts and filling the gaps in forecast horizons. However, as indicated in the discussion above of DRPS results, average performance over the full sample masks some differences over time. In estimates not reported in the interest of brevity, for a subsample of 2009-2016, tilting to the SPF bins improves the accuracy of model-based point and density forecasts as measured by RMSE and CRPS, respectively. The gains to tilting are on the order of 5 percent. These make the SV with tilting forecasts more accurate than the SV forecasts, whereas in the CONST case, tilting makes the forecasts as accurate as the SV benchmark, rather than less accurate in the purely model-based forecasts.

[Table 2 about here.]

5.5 Forecast performance without MDS assumption

So far, we have considered model estimates that treat the SPF point forecasts as optimal forecasts, and embody a martingale difference assumption regarding forecast updates. As described in Section 4.5, our model can also be extended to treat SPF predictions as martingales in only the model's Bayesian prior while allowing for VAR dynamics in the posterior estimates of the forecast update vector $\boldsymbol{\eta}_t$.

Table 3 reports the results of an out-of-sample evaluation of real-time forecasts generated by the MDS and non-MDS versions of the model. (Forecasts from the models used for this exercise do

not reflect entropic tilting.). We compare MDS and non-MDS specifications for both the SV model and the CONST model. As earlier, we consider point forecasts, evaluated in terms of RMSE, as well as density predictions, evaluated by the CRPS. From 1992Q1 onward, out-of-sample forecasts are generated for all quarterly horizons from $h = 0$ to 15, based on all available data since 1968Q4, by re-estimating the model at each forecast origin and simulating its predictive density.

As indicated in the upper panel of the table, the accuracy of point forecasts from the non-MDS specifications are generally very similar to those of their MDS-based counterparts — in both the full sample and the sample ending before the pandemic — with just a few exceptions. With SV, the RMSE ratios are 1.00 for most horizons, with a range of 0.98 to 1.02. With CONST (i.e., without SV), the RMSE ratios are also close to 1 in most cases, but in one instance hit 0.97 and in one other spike to 1.12 (in this case, without achieving statistical significance of the difference in MSE).

The CRPS results in the lower panel of the table indicate that the accuracy of density forecasts from the non-MDS specifications are generally similar to those of their MDS-based counterparts. However, for these density forecasts, with the SV model, the non-MDS specification registers a more consistent, small benefit in the pre-pandemic sample. Overall, in the collective results, the CRPS ratios are close to 1. For the CONST model, the non-MDS specification never improves on its MDS counterpart by more than 1 percent, and it is just as frequently 1 or 2 percent less accurate. But with SV included, in the pre-pandemic sample the non-MDS generalization lowers the CRPS by 1 to 3 percent at all horizons from 6 to 15 quarters. The small gains are much less consistent over the full sample.

Overall, these results suggest little benefit to generalizing our baseline model to allow departures from our MDS assumption. But they also show that there is little cost in forecast accuracy to doing so. Accordingly, our results are consistent with prior literature that, on one hand, reports considerable serial correlation in forecast updates of profession forecasters, like the SPF; see, for example, Coibion and Gorodnichenko (2015) and Coibion, Gorodnichenko, and Kamdar (2018). But, on the other hand, the literature has also found such deviations from SPF forecast efficiency to be episodic and hard to consistently exploit in real time, as noted, for example, by Croushore (2010), and more recently by Foerster and Matthes (2022), Hajdini and Kurmann (2022), and Mertens and Nason (2020). For our purposes, these results motivate our interest in centering estimates of the term structure of expectations and uncertainty around the SPF. We acknowledge,

however, that Bianchi, Ludvigson, and Ma (2022) have recently provided evidence that machine learning methods show more capability to improve on the accuracy of survey-based forecasts.

[Table 3 about here.]

5.6 Results for unemployment and inflation

Although to this point we have focused on GDP growth, the variable for which the SPF sample of annual forecasts is largest, we have also produced estimates for inflation (in GDP prices), and the unemployment rate. Many of the results we have described for GDP growth also apply to these variables. In the interest of brevity, in this section we examine forecasts of unemployment and inflation along two dimensions: (1) the impacts of entropic tilting to probability bins and (2) the impacts of generalizing the model so as to not impose the MDS assumption of the baseline. We focus on historical forecast accuracy, using the longest sample possible for each variable as well as a corresponding sample ending before the pandemic emerged. Table 4 provides RMSE and CRPS ratios for SV with tilting forecasts and non-MDS forecasts, both relative to the SV (without tilting) baseline.

Regarding the possible benefits of incorporating SPF bin forecasts through entropic tilting, the results for the unemployment rate and inflation in the GDP price index can be seen as consistent with the GDP growth results in that, while tilting does not offer consistent benefits, there can be periods in which it helps forecast accuracy. In results for the unemployment rate, the RMSE and CRPS ratios show that, for the 2009-2016 period, the tilted forecasts are notably less accurate than the entirely SV model-based forecasts, by as much as 18 percent for point forecasts and 23 percent for density forecasts. In this period, tilting is actually harmful to accuracy. But in the sample through 2022, tilting of unemployment forecasts achieves RMSE and CRPS ratios much closer to 1. Although tilting to the SPF's probability bin predictions was harmful until 2016, over the pandemic it instead improved the SV models's forecasts of unemployment. In results for inflation, the impacts of tilting are more similar across samples, reflecting the fact that, in the first year or so of the pandemic, inflation showed less variability than did measures of economic activity. For inflation, tilting slightly but consistently reduces the accuracy of the SV model's forecasts, both point and density, by 3 to 5 percent at most horizons.

As to the merits of the non-MDS specification as compared to the baseline SV model, the results for unemployment and inflation are very similar to those for GDP growth. The accuracy of point and density forecasts from the non-MDS specification is generally very similar to the MDS baseline — in both the full sample and the sample ending before the pandemic. The RMSE and CRPS ratios are close to 1. In the case of inflation, the non-MDS model offers a small gain (typically of 1-2 percent) for many horizons, for both point and density forecasts. But none of the gains appear to be statistically significant. For the unemployment rate, the score ratios are sometimes a little below 1 and other times a little above. So, for these variables as for GDP growth, the results suggest little benefit to generalizing our baseline model to allow departures from our MDS assumption.

[Table 4 about here.]

5.7 Practical applications using term structure of forecasts

The results above included fan charts for quarterly forecasts of GDP growth. As a practical matter, our model of the term structure of SPF forecasts can be used to produce SPF-consistent fan charts more directly analogous to those published by central banks. In some economies, central bank fan charts refer to four-quarter (year-over-year) growth rates, reported quarter by quarter. In the case of the Federal Reserve, since 2017 the FOMC has included in its Summary of Economic Projections (SEP) fan charts that report calendar year forecasts that are measured as Q4/Q4 growth rates of GDP and prices and the Q4 level of the unemployment rate. The width of the uncertainty bands in these fan charts is two times historical RMSEs that are reported in the SEP's Table 2, obtained as average RMSEs for a few different forecasts computed over 20 year rolling windows, as developed in Reifschneider and Tulip (2019). From 2007 through 2016, the SEP did not report fan charts, but it included historical RMSEs in its Table 2, with the intention of informing the FOMC participants' assessments of uncertainty as described below.[30]

To illustrate, Figures 9 and 10 provide SEP-style fan charts with calendar-year forecasts of GDP growth and the unemployment rate, for forecast origins of 2011Q1, 2012Q1, and 2014Q1. As detailed below, this period was associated with some changes in the FOMC's subjective assessments of uncertainty. Using estimates from our baseline SV model, the charts provide SPF-consistent point

[30]Historical RMSEs were provided as part of the SEP;s Table 2 since inception of the SEP format in October 2007; see `https://www.federalreserve.gov/monetarypolicy/files/fomcminutes20071031.pdf`.

forecasts and 68 percent uncertainty bands.[31] Across these origins, the SPF's point forecasts of growth were fairly flat across horizons, with uncertainty gradually rising with the horizon. As we will detail below, our estimates of uncertainty (the width of the bands in these charts) were well below the estimates used in the SEP's fan charts. The SPF's point forecasts of the unemployment rate declined significantly across horizons in 2011Q1 and 2012Q1 and edged down more modestly in 2014:Q1, in all cases with uncertainty gradually rising with the horizon.

Figure 10 directly compares over time uncertainty as estimated from our model (with the width of 68 percent bands) to uncertainty measured as the SEP currently does, as two times the historical RMSEs published in the SEP's Table 2. For each year from 2008 through 2021, we report estimates for SPF and SEP forecasts published in the first quarter.[32] As might be expected, our estimates show more variation over time than do the SEP measures, because the model includes SV, whereas the SEP measures treat forecast error variances as being constant over 20 year windows. Following the volatility induced by the Great Recession, our estimates of uncertainty rise relatively quickly and sharply; the SEP measures eventually reflect effects of the large forecast errors of the recession, but rise relatively slowly and by less than the SV estimates. But our SV-based estimates fell quickly. In the 2011Q1, 2012Q1, and 2014Q1 examples, SPF-consistent forecasts from our model imply noticeably less uncertainty in the outlook for GDP growth than the SEP-based approach. The same pattern applies to the unemployment rate examples, except that, in 2011Q1, our estimates of uncertainty were relatively similar to the SEP-based measures. From 2015 to 2019, both our estimates and the SEP estimates are relatively stable, with the latter continuing to exceed the former. Following the volatility induced by the outbreak of the pandemic in 2020, our SV-based estimates of uncertainty rose sharply in 2021Q1, to exceed the SEP estimates.

In addition to providing fan charts and historical RMSEs to quantify uncertainty around the outlook, the SEP includes measures of FOMC participants' subjective assessments of uncertainty. In the current SEP, a note in the fan charts explains: "Because current conditions may differ from those that prevailed, on average, over the previous 20 years, the width and shape of the confidence interval estimated on the basis of the historical forecast errors may not reflect FOMC participants'

[31] Q1 SPF forecasts are published in mid-February of each year. What we treat as the Q1 SEPs were published in late January (2011 and 2012) or mid-March (2014).

[32] The available time series for the SEP begins in October 2007, making the first available Q1 observation be 2008. The charts omit SEP observations for 2020 because, in March 2020, in the aftermath of the pandemic's outbreak, the FOMC did not publish an SEP.

current assessments of the uncertainty and risks around their projections; these current assessments are summarized in the lower panels. Generally speaking, participants who judge the uncertainty about their projections as "broadly similar" to the average levels of the past 20 years would view the width of the confidence interval shown in the historical fan chart as largely consistent with their assessments of the uncertainty about their projections."[33] The subjective assessments of participants are collected from responses to the following survey question included in each SEP: "Please indicate your judgment of the uncertainty attached to your projections relative to the levels of uncertainty over the past 20 years," with answer options of "broadly similar," "lower," or "higher."[34] Since December 2020, the SEP has included in Figure 4.D a chart of a diffusion index that summarizes the Committee's subjective assessments of uncertainty.

As this makes clear, the SEP's subjective assessments of uncertainty may well differ from historical quantitative assessments. The period from 2011 to 2014 provides such an example. In our model estimates, uncertainty was relatively stable from 2011 and 2012 to 2014. While the levels of estimates based on the SEP's current treatment of fan charts were generally higher than our estimates, the SEP-type measures also showed little change across these three years, except that GDP growth uncertainty rose from 2012 to 2014. In contrast, the SEP's subjective assessments show some relatively sizable changes over the period. From 2012 to 2014, the subjective uncertainty assessments around growth and unemployment shifted from 'elevated' to 'normal', reflected in a drop in the diffusion index.

[Figure 9 about here.]

[Figure 10 about here.]

6 Conclusion

For application to professional forecasts such as the SPF, this paper develops a model that permits the estimation of a term structure of both expectations and forecast uncertainty. In broad terms, our model can be seen as a multivariate unobserved component model with stochastic volatility.

[33] See the note to Figure 4.A in https://www.federalreserve.gov/monetarypolicy/files/fomcprojtabl20220316.pdf.

[34] The question and other SEP information is provided in https://www.federalreserve.gov/monetarypolicy/guide-to-the-summary-of-economic-projections.htm.

Our approach exactly replicates a given data set of predictions from the SPF or a similar forecast source without measurement error. Extending some previous work, our model includes not only fixed-horizon forecasts but also fixed-event forecasts, including at horizons beyond the fixed-horizon maximum, and it accommodates variation over time in the available horizons of fixed-event forecasts. We bring the density forecast information contained in the SPF's probability bins for fixed-event forecasts to bear on the term structure of uncertainty through entropic tilting, using tilting to adjust the model-based predictive distribution so that the tilted distribution's implied probability bins match up to those from the SPF (while minimizing a distance criterion). This application of entropic tilting may help improve the uncertainty estimate and density forecast accuracy, particularly at horizons longer than the quarters covered in a given published SPF's quarterly projections, and it allows us to treat the SPF's subjective point and density forecasts commensurately.

Our empirical results first establish that our model yields quarterly forecasts that perfectly match and interpolate through the annual fixed-event point forecasts; sometimes the interpolation is quite smooth, but in other instances, it shows some variation of quarterly forecasts around the annual predictions. Examples of quarterly fan charts of point forecasts and uncertainty bands around those forecasts show that the model's estimates of uncertainty vary over time and generally rise with the forecast horizon. Our assessment of the efficacy of using entropic tilting to incorporate information in the SPF's subjective probability bins for fixed-event annual forecasts, reveals that, over the full sample, the model's point and density forecasts based on SPF point forecasts are relatively good; on average, tilting to the SPF's probability bins does not offer consistent benefits. That said, there can be periods in which tilting to the bin information helps forecast accuracy — for both point and density forecasts. Finally, we apply our approach to constructing fan charts with SPF-based calendar year (four-quarter growth rates) forecasts like those published by the FOMC. In recent years, our model applied to SPF forecasts yields estimates of uncertainty around GDP growth and unemployment forecasts that are lower than those implied by the historical forecast RMSEs that underlay the FOMC's fan charts. Our approach could be used in real time with each new release of the SPF to produce updated fan charts of point forecasts and estimates of forecast uncertainty.

Appendix A Implementation of Entropic Tilting

Entropic tilting described in Section 4.8 does not admit a closed-form solution in general. Instead we resort to numerical methods to solve the underlying constrained minimization problem with details described here.

Recall that at a given forecast origin t, each model (e.g. SV or CONST) delivers a set of predictive draws of quarterly growth rates (in the case of real GDP and the GDP price index) or quarterly levels (in the case of the unemployment rate), which we transformed to *calendar-year* objects $\{\boldsymbol{X}_t^s\}_{s=1}^S$, as explained in the main text. The dimension of $\boldsymbol{X}_t^s$ is given by the number of calendar-year horizons at which we aim to match the SPF histograms. Next, given the SPF bin structure, to each MCMC draw we assign the value of either 0 or 1, depending on whether the specific MCMC draw is outside (0) or inside (1) each bin. Staying with the example in Section 4.8, in the 2021Q4 SPF round the predictive draws of the unemployment rate for each of the target years 2022, 2023 and 2024 fell into one of the $B = 9$ bins, and this we recorded in the $(3 \cdot 9 \times S)$ matrix, whose columns are associated with the draws, and rows correspond to the bins at each calendar-year horizon. The resulting array is multiplied by 100, so relative frequencies can be interpreted as percentage terms. The final ingredient to entropic tilting is the vector of probabilities (again, in percentage terms) corresponding to each bin, interpreted as moment conditions, given by $\bar{\boldsymbol{g}}_t \equiv \bar{\boldsymbol{p}}_t$.

The solution to the minimization problem in equation (21) is not available in closed form, and we used an unconstrained optimizer.[35] In particular, we set the 0 vector as the starting point, which corresponds to using the MCMC output's implied histograms. To aide the computations, our implementation relies on the available closed-form expressions of gradient and the Hessian of the objective. We used MATLAB's `trust-region` algorithm with $1,500$ maximum iterations, keeping the rest of the settings at their defaults. The numerical optimizer was considered to have failed if one of the following three conditions was met: a) numerical underflow/overflow was encountered in computing the weights $\widetilde{w}_s^*$, b) the minimizer did not converge, c) the proposed solution led to moment conditions whose maximum absolute difference against the SPF moment conditions was larger than one percentage point. In any of these cases, inspired by Krüger, Clark, and Ravazzolo (2017), the minimization was restarted after adding a penalty term to the objective function, which

[35] We used the function `fminunc` provided by MATLAB's optimization toolbox.

we defined as the product of three terms: (i) the value of the objective function at the initial (but not optimal) solution returned by the optimizer, (ii) a shrinkage term, and (iii) the square of the Euclidean norm of the γ vector. This procedure was repeated for various values of the shrinkage term (specified on a grid ranging from 10^{-10} to 10, in 20 equally spaced steps) until the optimizer converged and we obtained valid weights, which implied probabilities within the 1 percent tolerance we set.

References

[1] 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.

[2] Arias, Jonas E., Juan F. Rubio-Ramirez, and Minchul Shin (2022), "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Journal of Econometrics*, forthcoming.

[3] Aruoba, S. Boragan (2020), "Term structures of inflation expectations and real interest rates," *Journal of Business & Economic Statistics*, 38, 542–553, `https://doi.org/10.1080/07350015.2018.152`.

[4] Bassetti, Federico, Roberto Casarin, and Marco Del Negro (2022), "Inference on probabilistic surveys in macroeconomics with an application to the evolution of uncertainty in the Survey of Professional Forecasters during the COVID pandemic," in *Handbook of Economic Expectations* eds. by W. van der Klaauw, G. Topa, and R. Bachmann: Elsevier.

[5] Beveridge, Stephen, and Charles R. Nelson (1981), "A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the "Business Cycle"," *Journal of Monetary Economics*, 7, 151–174.

[6] 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`.

[7] Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2011), "Scoring rules and survey density forecasts," *International Journal of Forecasting*, 27, 379–393, https://doi.org/10.1016/j.ijforecast.2010.04.003.

[8] Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2016), "Common drifting volatility in large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390, https://doi.org/10.1080/07350015.2015.1040116.

[9] Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2021), "Addressing COVID-19 outliers in BVARs with stochastic volatility," *Review of Economics and Statistics*, forthcoming.

[10] 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.

[11] Chan, Joshua C.C., and Ivan Jeliazkov (2009), "Efficient simulation and integrated likelihood estimation in state space models," *International Journal of Mathematical Modelling and Numerical Optimization*, 1, 101–120, https://doi.org/10.1504/IJMMNO.2009.030090.

[12] Chib, Siddhartha, and Ivan Jeliazkov (2006), "Inference in semiparametric dynamic models for binary longitudinal data," *Journal of the American Statistical Association*, 101, 685–700, https://doi.org/10.1198/016214505000000871.

[13] Clark, Todd E. (2011), "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business and Economic Statistics*, 29, 327–341, https://doi.org/10.1198/jbes.2010.09248.

[14] 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.

[15] Clark, Todd E., and Francesco Ravazzolo (2015), "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30, 551–575, https://doi.org/10.1002/jae.2379.

[16] Clements, Michael P. (2018), "Are macroeconomic density forecasts informative?" *International Journal of Forecasting*, 34, 181–198, https://doi.org/10.1016/j.ijforecast.2017.

[17] Clements, Michael P., and Ana Beatriz Galvão (2017), "Model and survey estimates of the term structure of US macroeconomic uncertainty," *International Journal of Forecasting*, 33, 591–604, https://doi.org/10.1016/j.ijforecast.2017.01.004.

[18] Cogley, Timothy, Sergei Morozov, and Thomas J. Sargent (2005), "Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system," *Journal of Economic Dynamics and Control*, 29, 1893–1925, https://doi.org/10.1016/j.jedc.2005.06.005.

[19] Coibion, Olivier, and Yuriy Gorodnichenko (2015), "Information rigidity and the expectations formation process: A simple framework and new facts," *American Economic Review*, 105, 2644–78, https://doi.org/10.1257/aer.20110306.

[20] Coibion, Olivier, Yuriy Gorodnichenko, and Rupal Kamdar (2018), "The formation of expectations, inflation, and the Phillips curve," *Journal of Economic Literature*, 56, 1447–1491, https://doi.org/10.1257/jel.20171300.

[21] Croushore, Dean (2010), "An Evaluation of Inflation Forecasts from Surveys Using Real-Time Data," *The B.E. Journal of Macroeconomics*, 10, 1–32.

[22] Croushore, Dean, and Tom Stark (2001), "A real-time data set for macroeconomists," *Journal of Econometrics*, 105, 111–130, https://doi.org/10.1016/S0304-4076(01)00072-0.

[23] Crump, Richard K., Stefano Eusepi, Emanuel Moench, and Bruce Preston (2021), "The term structure of expectations," Staff Reports 992, Federal Reserve Bank of New York.

[24] D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013), "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101, https://doi.org/10.1002/jae.1257.

[25] Diebold, Francis X., Anthony S. Tay, and Kenneth F. Wallis (1999), "Evaluating density forecasts of inflation: The survey of professional forecasters," in *Cointegration, Causality, and*

Forecasting: A Festschrift in Honour of Clive WJ Granger eds. by R. Engle, and H. White: Oxford University Press, 76–90.

[26] Dieppe, Alistair, Romain Legrand, and Björn van Roye (2016), "The BEAR Toolbox," Working Paper No. 1934, European Central Bank.

[27] Doh, Taeyoung, and Andrew Lee Smith (2021), "Reconciling VAR-based forecasts with survey forecasts," Research Working Paper RWP 18-13, Federal Reserve Bank of Kansas City, https://doi.org/10.18651/RWP2018-13.

[28] Dovern, Jonas, Ulrich Fritsche, and Jiri Slacalek (2012), "Disagreement among forecasters in G7 countries," *Review of Economics and Statistics*, 94, 1081–1096, https://doi.org/10.1162/REST_a_00207.

[29] Durbin, J., and S.J. Koopman (2002), "A simple and efficient simulation smoother for state space time series analysis," *Biometrika*, 89, 603–615, https://doi.org/10.1093/biomet/89.3.603.

[30] 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.

[31] Foerster, Andrew, and Christian Matthes (2022), "Learning about regime change," *International Economic Review*, forthcoming, https://doi.org/10.1111/iere.12585.

[32] 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.

[33] Galvão, Ana Beatriz, Anthony Garratt, and James Mitchell (2021), "Does judgment improve macroeconomic density forecasts?" *International Journal of Forecasting*, 37, 1247–1260, https://doi.org/10.1016/j.ijforecast.2021.02.007.

[34] 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.

[35] Ganics, Gergely, Barbara Rossi, and Tatevik Sekhposyan (2021), "From fixed-event to fixed-horizon density forecasts: Obtaining measures of multi-horizon uncertainty from survey density forecasts," *Journal of Money, Credit, and Banking*, forthcoming.

[36] Glas, Alexander, and Matthias Hartmann (2021), "Uncertainty measures from partially rounded probabilistic forecast surveys," *Quantitative Economics*, forthcoming.

[37] 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.

[38] Hajdini, Ina, and André Kurmann (2022), "Predictable forecast errors in full-information rational expectations models with regime shifts," *mimeo*.

[39] Hepenstrick, Christian, and Jason Blunier (2022), "What were they thinking? Estimating the quarterly forecasts underlying annual growth projections," Working Paper 2022-05, Swiss National Bank.

[40] Jo, Soojin, and Rodrigo Sekkel (2019), "Macroeconomic uncertainty through the lens of professional forecasters," *Journal of Business & Economic Statistics*, 37, 436–446, https://doi.org/10.1080/07350015.2017.1356729.

[41] 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.

[42] 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.

[43] 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.

[44] 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.

[45] Mertens, Elmar (2022), "A precision-based sampler for unobserved component models without measurement error," *mimeo*.

[46] 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.

[47] Omori, Yasuhiro, Siddhartha Chib, Neil Shephard, and Jouchi Nakajima (2007), "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425–449, https://doi.org/10.1016/j.jeconom.2006.07.008.

[48] 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.

[49] ——— (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.

[50] 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.

[51] Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2005), "Forecasting using relative entropy," *Journal of Money, Credit and Banking*, 37, p. 20, https://doi.org/10.1353/jmcb.2005.0034.

[52] Tallman, Ellis W., and Saeed Zaman (2020), "Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy," *International Journal of Forecasting*, 36, 373–398, https://doi.org/10.1016/j.ijforecast.2019.04.024.

[53] 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: Availability of SPF predictions

		Fixed-horizons	Fixed-event calendar years		
Variable	Mnemonic	Quarters 0 – 4	next	2-year	3-year
Panel A: Point forecasts					
Real GDP	RGDP	1968Q4	1981Q3	2009Q2	2009Q2
Unemployment rate	UNRATE	1968Q4	1981Q3	2009Q2	2009Q2
GDP price index	PGDP	1968Q4	1981Q3	NA	NA
Panel B: Histograms					
Real GDP growth	PRGDP	NA	1981Q3	2009Q2	2009Q2
Unemployment rate	PRUNEMP	NA	2009Q2	2009Q2	2009Q2
GDP price index inflation	PRPGDP	NA	1981Q3	NA	NA

Note: The table reports the first quarters since when SPF predictions are available in our data set for the stated variables and horizons. NA stands for not available. Prior to 1992, RGDP and PRGDP correspond to real GNP, while PGDP and PRPGDP correspond to the GNP implicit deflator. The SPF solicits point forecasts for RGDP and PGDP in levels, which we convert to continuously compounded growth rates. The SPF also provides current year predictions (point and density forecasts) which are, however, disregarded in our analysis due to overlap with the fixed-horizon predictions and narrowing current-year densities towards the end of each calendar year.

Table 2: Accuracy of GDP growth forecasts with and without entropic tilting

			Relative to SV					
	SV		SV w/ET		CONST		CONST w/ET	
h	92–22	92–16	92–22	92–16	92–22	92–16	92–22	92–16
PANEL A: RMSE								
0	2.04	1.70	1.01	1.01*	1.00	1.00	1.01**	1.01*
1	4.52	1.97	1.00	1.01	1.00	1.00	1.00	1.01
2	4.88	2.11	1.00	1.00	1.00	1.00	1.00	1.00
3	4.96	2.24	1.00	0.99	1.00	1.00	1.00	0.99
4	5.02	2.29	1.00	0.99	1.00	1.00	1.00	0.99
5	5.07	2.34	1.00	0.98	1.00	1.00	1.00	0.99
6	5.07	2.29	1.00	0.99	1.00**	1.01	1.01	1.01
7	5.08	2.32	1.00	0.99	1.00	1.01	1.00	1.00
8	5.12	2.29	1.00	0.99	1.00	1.02	1.00	1.01
9	5.10	2.24	1.00	0.99	1.00	1.01	1.00	1.00
10	5.11	2.20	1.00	1.00	1.00	1.02	1.00	1.01
11	5.16	2.21	1.00	0.99	1.00	1.02*	1.00	1.01
12	5.18	2.20	1.00	1.00	1.00	1.02	1.00	1.01
13	5.20	2.32	1.00	1.01*	1.00	1.02	1.00	1.02
14	5.23	4.61	1.00	1.00	1.00	1.00	1.01	1.00
15	5.26	5.33	1.00	1.00	1.00	1.00	1.00	1.00
PANEL B: CRPS								
0	1.07	0.98	1.00	1.00	0.99	0.99	1.00	0.99
1	1.51	1.09	1.00	1.00	1.04	1.01	1.04*	1.02
2	1.68	1.14	0.99	1.00	1.00	1.02	1.01	1.02
3	1.73	1.20	0.99	0.99	1.01	1.02	1.01	1.01
4	1.77	1.20	0.99	0.99	1.01	1.03*	1.01	1.01
5	1.85	1.25	0.99	0.98	1.01	1.04*	1.01	1.01
6	1.82	1.23	0.99	0.99	1.03	1.06**	1.03	1.04
7	1.85	1.25	1.00	0.99	1.03	1.06**	1.03	1.04
8	1.86	1.24	1.00	0.99	1.03	1.06**	1.03	1.04
9	1.83	1.21	1.00	0.99	1.03	1.06**	1.03	1.05
10	1.83	1.20	1.00	0.99	1.03**	1.07***	1.04**	1.06**
11	1.85	1.20	0.99	0.98	1.04**	1.07***	1.03	1.05*
12	1.86	1.20	0.99	0.99	1.04**	1.07***	1.03**	1.06**
13	1.87	1.25	1.00	1.00	1.04***	1.07***	1.04***	1.06***
14	1.88	1.63	1.00	1.00	1.04**	1.05***	1.04**	1.05**
15	1.90	1.88	0.99*	0.99	1.04**	1.04**	1.04**	1.04**

Note: Forecasts for quarterly GDP growth h steps ahead. Ratios of RMSE and CRPS use SV model in denominator. Significance assessed by Diebold-Mariano test using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively. "92–22" and "92–16" refer to subsamples extending from 1992Q1 until 2022Q2 and 2016Q4, respectively (using data for realized values as far as available in 2022Q2).

Table 3: Forecast accuracy for GDP growth without MDS assumption

	SV		CONST	
h	92–22	92–16	92–22	92–16
PANEL A: RMSE				
0	1.00	1.00	1.12	1.01
1	1.02	1.00	1.06	1.01
2	1.00	0.98*	1.00	0.97**
3	1.00	1.00	1.00	0.99
4	1.00	1.00	1.00	0.99
5	1.00	1.00	1.00	1.01
6	1.00	1.00	1.00	1.02
7	1.00	1.00	1.00	0.99
8	1.00	1.01	1.00	1.00
9	1.00	1.01	1.01	1.01
10	1.00	1.01	1.01	1.02
11	1.00	1.01	1.00	1.02
12	1.00	1.00	1.00	1.02
13	1.00	1.00	1.01	1.01
14	1.00	1.00	1.00	1.00
15	1.00	1.00	1.00	1.00
PANEL B: CRPS				
0	1.01	1.01	1.08	1.01
1	1.02	1.01	1.04	1.01
2	1.01	0.99*	1.00	0.99
3	1.02	1.01	1.00	1.00
4	1.02	1.01	1.01	1.01
5	1.01	1.00	1.01	1.00
6	1.00	0.99	1.01	1.01
7	1.00	0.99	0.99	0.99
8	1.00	0.99	0.99	0.99
9	1.00	0.99	1.01	1.01
10	0.99	0.98	1.01	1.01
11	1.00	0.98	1.01	1.02
12	0.99	0.97**	1.01	1.01
13	0.99	0.97**	1.00	1.00
14	0.99	0.98	1.00	1.01
15	0.98*	0.98**	1.00	1.00

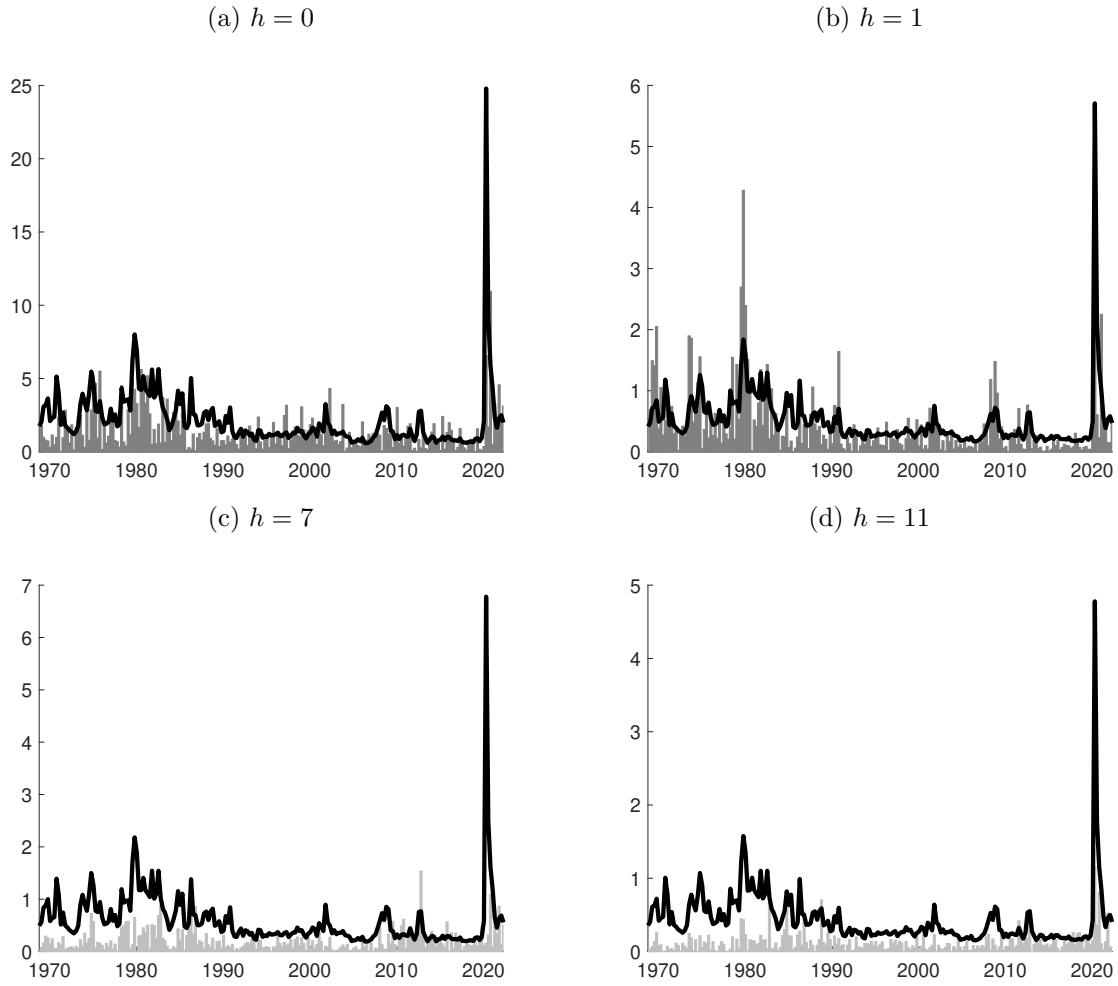
Note: Comparison of forecasts with and without MDS assumption for quarterly GDP growth h steps ahead. Relative RMSE and CRPS of non-MDS forecasts with corresponding MDS statistics in the denominator.Significance assessed by Diebold-Mariano test using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively. "92–22" and "92–16" refer to subsamples beginning in 1992Q1 and extending until 2022Q2, and 2016Q4, respectively, (using data for realized values as far as available in 2022Q2).

Table 4: Forecast performances for additional variables

	PGDP				UNRATE			
	SV w/ET		non-MDS		SV w/ET		non-MDS	
h	92–22	92–16	92–22	92–16	09–22	09–16	09–22	09–16
PANEL A: RMSE								
0	1.01*	1.01	0.99	1.01	1.00	1.02	0.95	0.96
1	1.01	1.03*	1.00	1.00	1.00	1.02	0.99	0.96
2	1.02**	1.04*	0.99	0.99	0.99	1.05	1.01	0.99
3	1.02*	1.05*	0.99	0.99	0.99	1.08*	1.01	1.01
4	1.03**	1.05**	0.99	0.98	0.99	1.12**	1.01	1.01
5	1.02	1.03	0.99	0.95	0.99	1.13***	1.01	1.00
6	1.02**	1.04**	1.01	1.02	0.98	1.15***	1.02	0.99
7	1.02**	1.03*	1.02	1.03	0.97	1.17***	1.01	1.00
8	–	–	–	–	0.98	1.18***	1.00	1.01
9	–	–	–	–	1.00	1.18***	1.00	1.00
10	–	–	–	–	1.00	1.19***	1.00	1.00
11	–	–	–	–	1.00	1.20***	1.00	1.00
12	–	–	–	–	1.02	1.20***	1.00	1.00
13	–	–	–	–	1.02	1.18***	1.00	0.99
14	–	–	–	–	1.03	1.06***	1.00	1.00
15	–	–	–	–	1.03	1.04**	1.00	1.00
PANEL B: CRPS								
0	1.01	1.01	1.00	1.00	1.03*	1.02*	0.96	0.95
1	1.02*	1.02*	1.00	1.00	1.02*	1.03	1.00	0.97
2	1.03**	1.03**	0.99	0.98	1.00	1.05	1.01	0.99
3	1.03**	1.04**	0.99	0.98	0.98	1.07	1.01	1.00
4	1.03**	1.04**	0.99	0.99	0.97	1.09*	1.01	1.01
5	1.02*	1.02*	0.98	0.95	0.96	1.12**	1.01	1.00
6	1.02**	1.03**	1.01	1.00	0.95	1.13**	1.01	0.99
7	1.02**	1.03**	1.02	1.01	0.96	1.15**	1.00	1.00
8	–	–	–	–	1.02	1.17**	1.00	1.01
9	–	–	–	–	1.04	1.19***	1.00	1.00
10	–	–	–	–	1.05	1.21***	1.00	1.00
11	–	–	–	–	1.07	1.23***	1.00	1.00
12	–	–	–	–	1.08	1.23***	1.00	1.00
13	–	–	–	–	1.09	1.22***	1.00	0.99
14	–	–	–	–	1.09	1.15***	1.00	1.00
15	–	–	–	–	1.08	1.11***	1.00	1.00

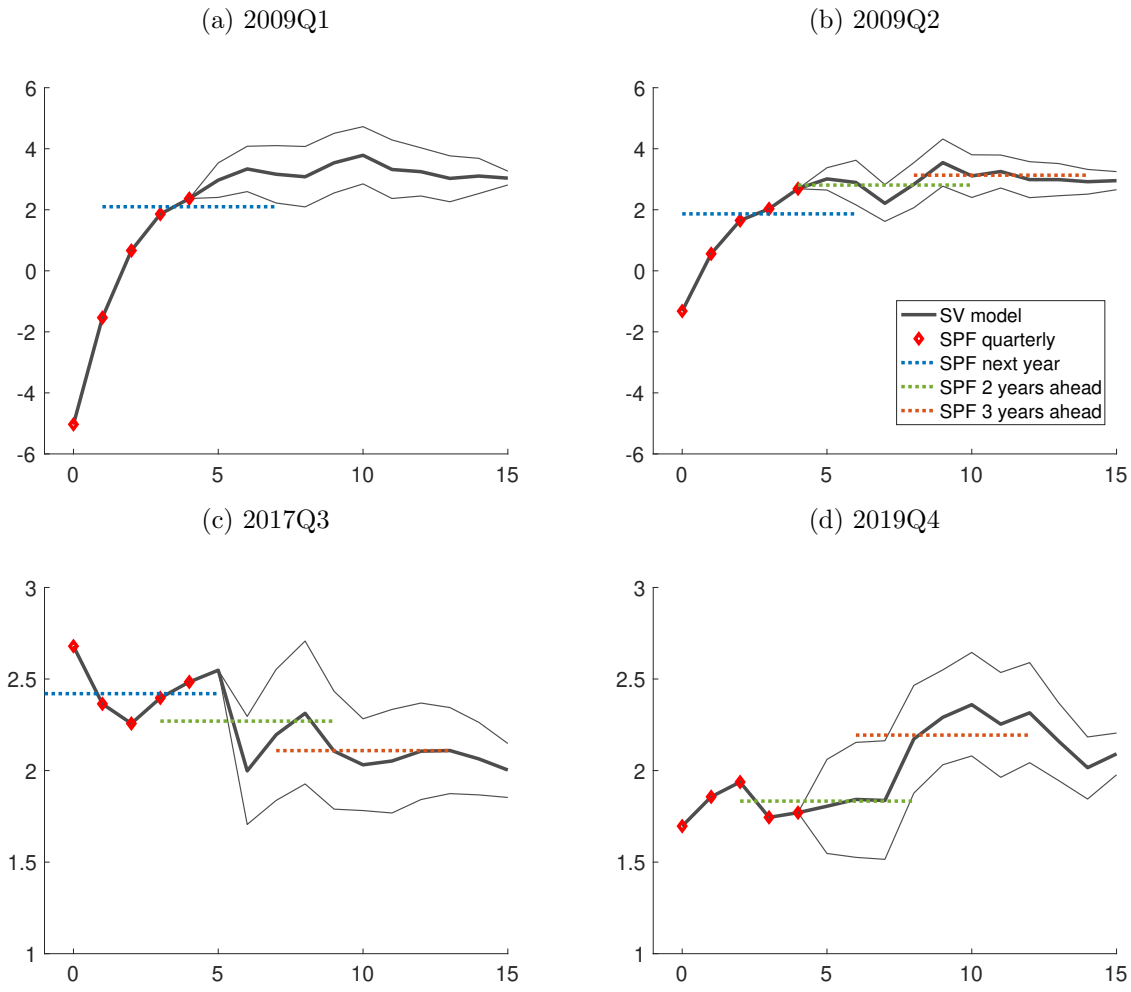
Note: RMSE and CRPS of each model relative to SV (in denominator). Significance assessed by Diebold-Mariano test using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively. Evaluation windows extend from 1992Q1 and 2009Q2 until 2022Q2 and 2016Q4, respectively, and as indicated by the column label "92–22," "09–22," "09–22," and "09–16." (using data for realized values as far as available by 2022Q2).

Figure 1: Stochastic volatility estimates for GDP growth



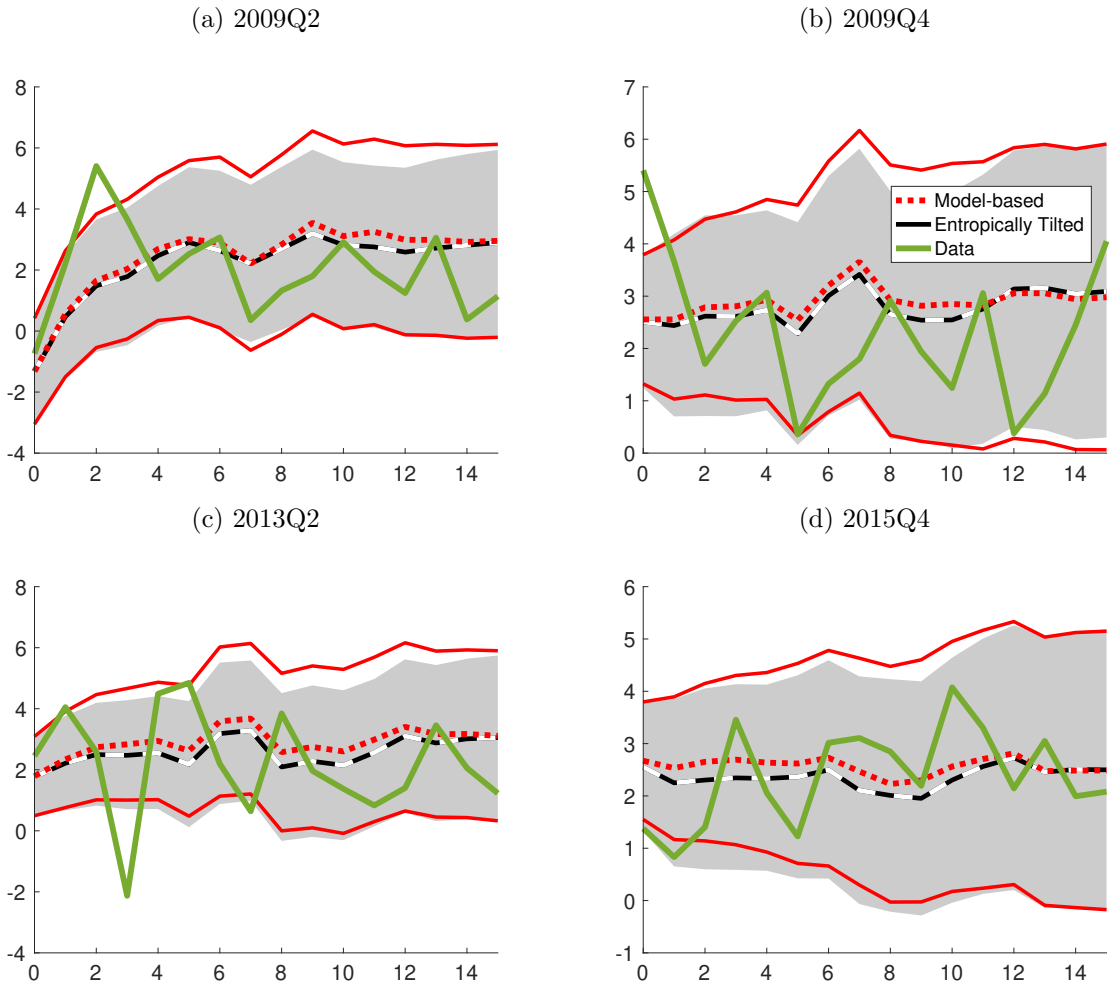
Notes: Stochastic volatility estimates for forecast updates, $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$, of GDP growth at horizons $h = 0, 1, 7, 11$. Grey bars report the absolute values of the forecast errors at each horizon; darker grey bars for $h = 0$ and $h = 1$ indicate that forecast updates for those horizons are directly observed, whereas lighter grey bars for $h = 7$ and $h = 11$ reflect posterior medians obtained from our baseline SV model applied to the full data sample through 2022Q2. Black lines report (smoothed) volatility estimates obtained with the full sample of data (through 2022Q2).

Figure 2: Term structures of GDP growth expectations



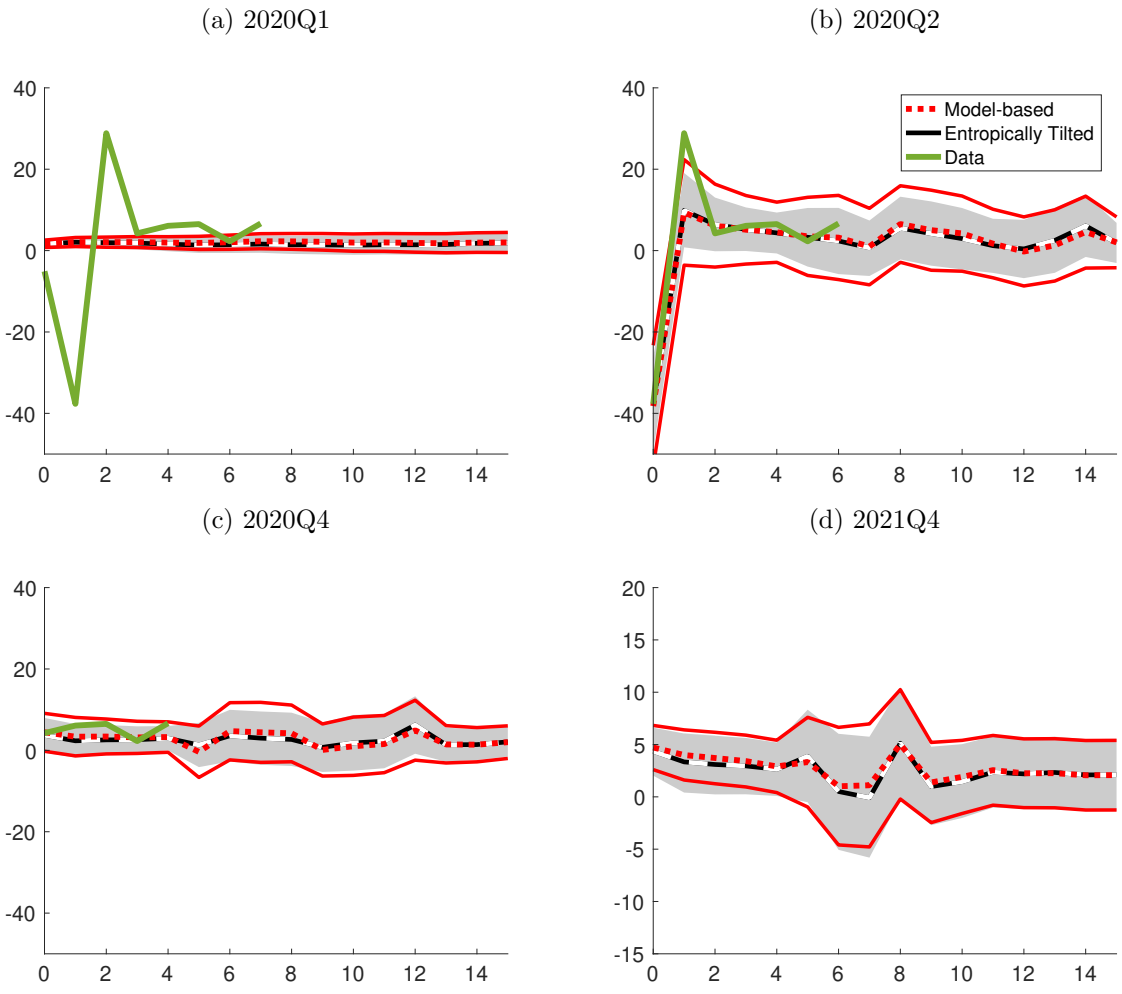
Notes: Term structure of GDP growth expectations, selected forecast origins. Real-time estimates. Red diamonds give the available quarterly fixed-horizon forecasts from SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the quarters included in the tent-shaped linear mapping from quarterly to annual-average changes). Horizons (in quarters) are indicated on the horizontal axis.

Figure 3: Forecast fan charts for GDP growth



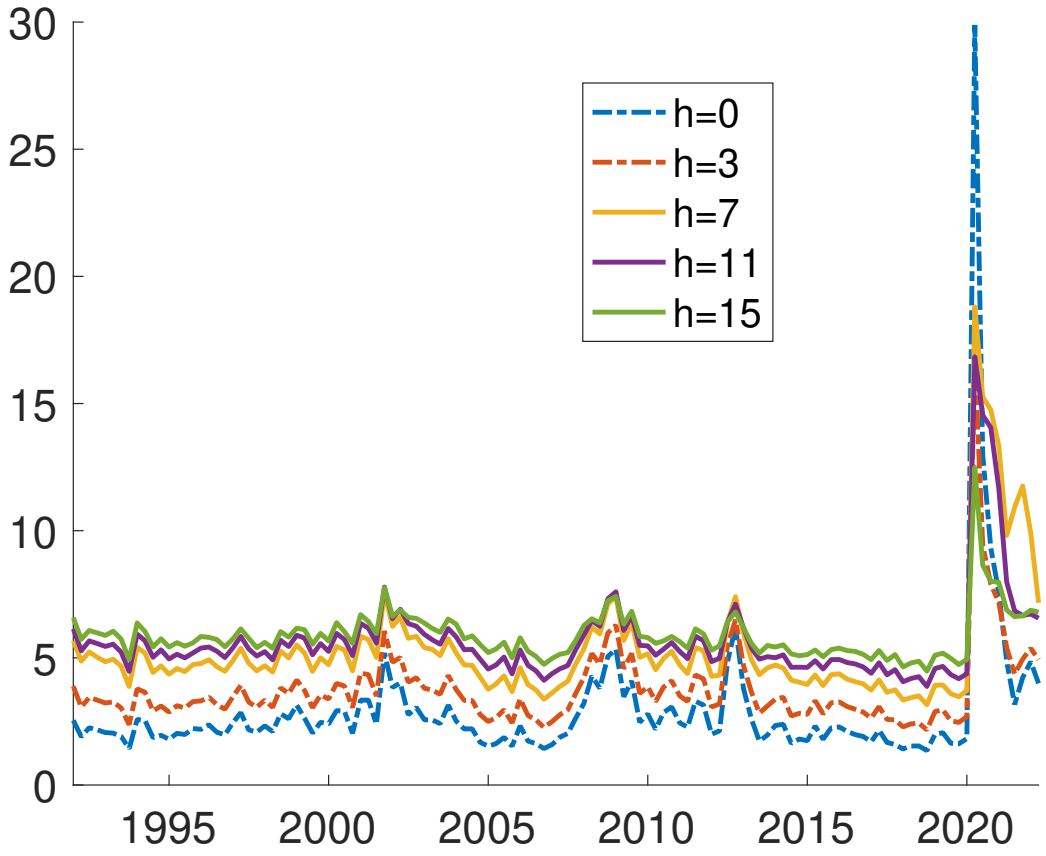
Notes: Forecast fan charts for GDP growth from the baseline SV model. Model-based (red), entropically tilted (black and grey), with 68% uncertainty bands each, and realized values (green).

Figure 4: Forecast fan charts for GDP growth since 2020



Notes: Forecast fan charts for GDP growth since 2020 from the baseline SV model. Model-based (red), entropically tilted (black and grey), with 68% uncertainty bands each, and realized values (green).

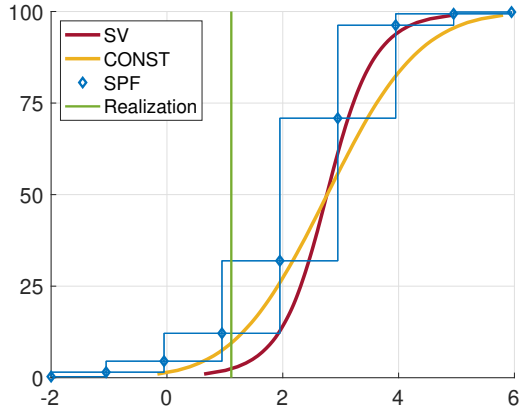
Figure 5: Term structures of GDP growth uncertainty



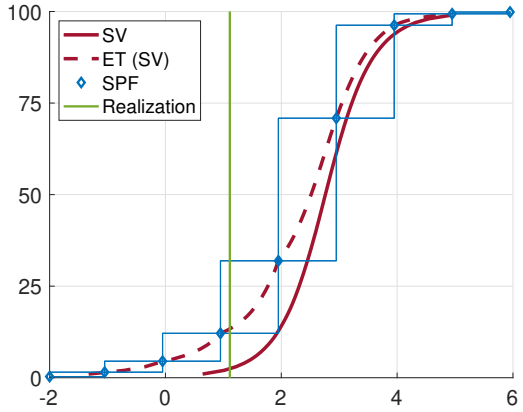
Notes: Term structures of GDP growth uncertainty, measured by the width of 68% bands of predictive real-time densities from the baseline SV model.

Figure 6: Cumulative density functions for GDP growth with and without tilting, in 2007Q3

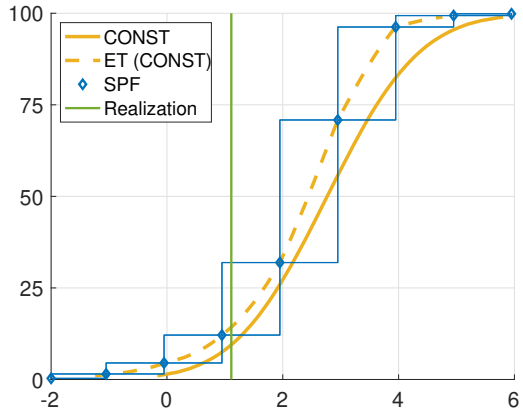
(a) SV vs CONST



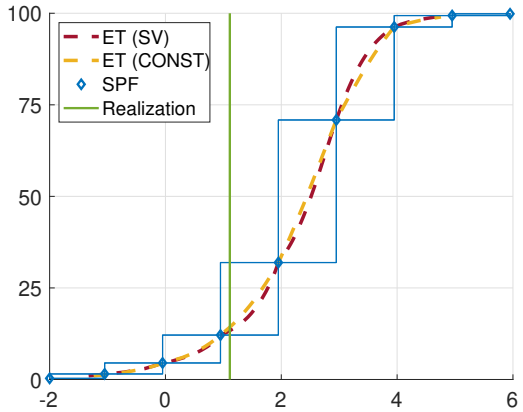
(b) SV vs SV w/ET



(c) CONST vs CONST w/ET



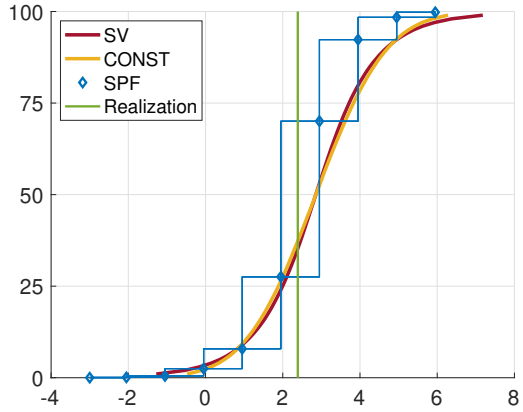
(d) SV w/ET vs CONST w/ET



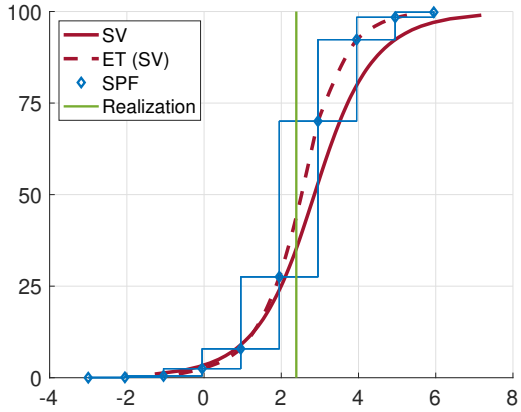
Notes: Cumulative densities obtained from our SV model (with and without entropic tilting), as well as the SPF histograms. The SPF histograms report only discrete probabilities (marked by diamonds), and leave open the precise shape of the underlying density for values in between (as marked by the boxes that are demarcated with blue dotted lines). Probabilities on the y-axis in percentage points, and values for annualized GDP growth in percent on the x-axis.

Figure 7: Cumulative density functions for GDP growth with and without tilting, in 2013Q1

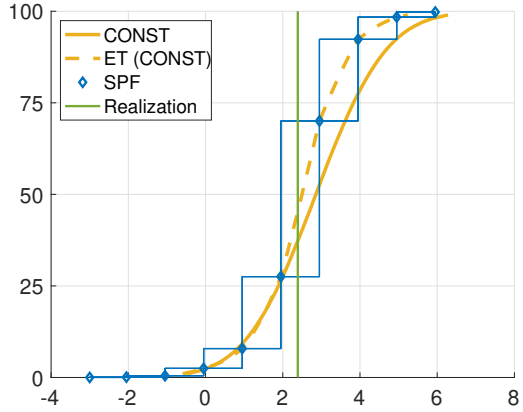
(a) SV vs CONST



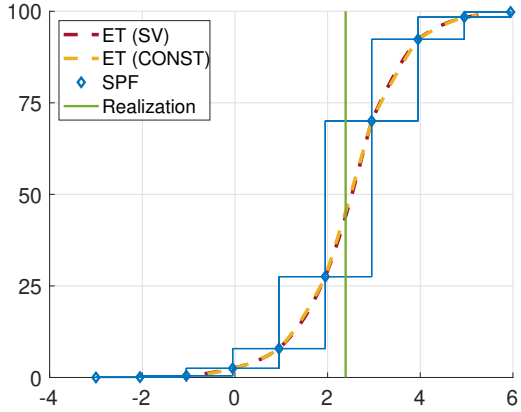
(b) SV vs SV w/ET



(c) CONST vs CONST w/ET

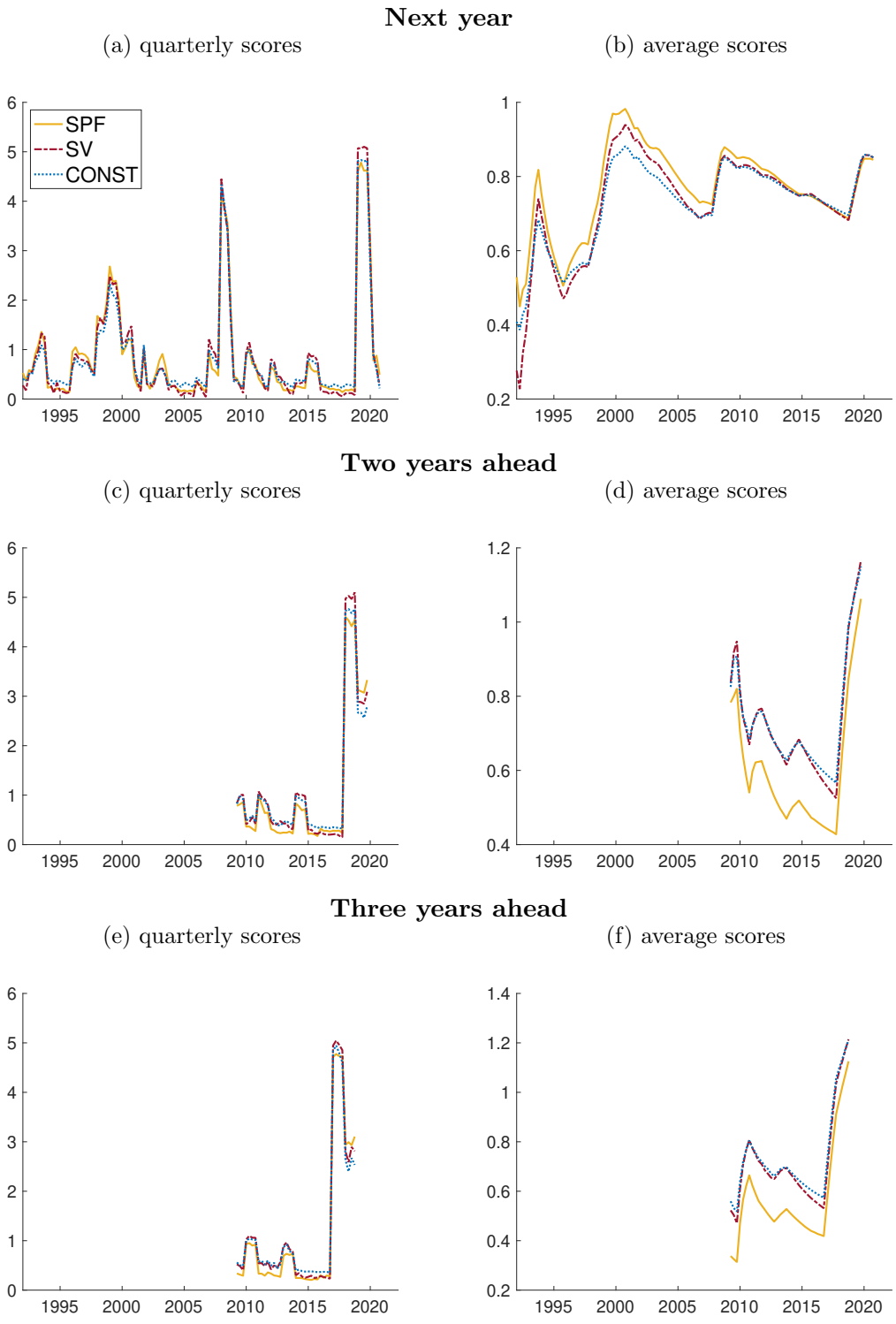


(d) SV w/ET vs CONST w/ET



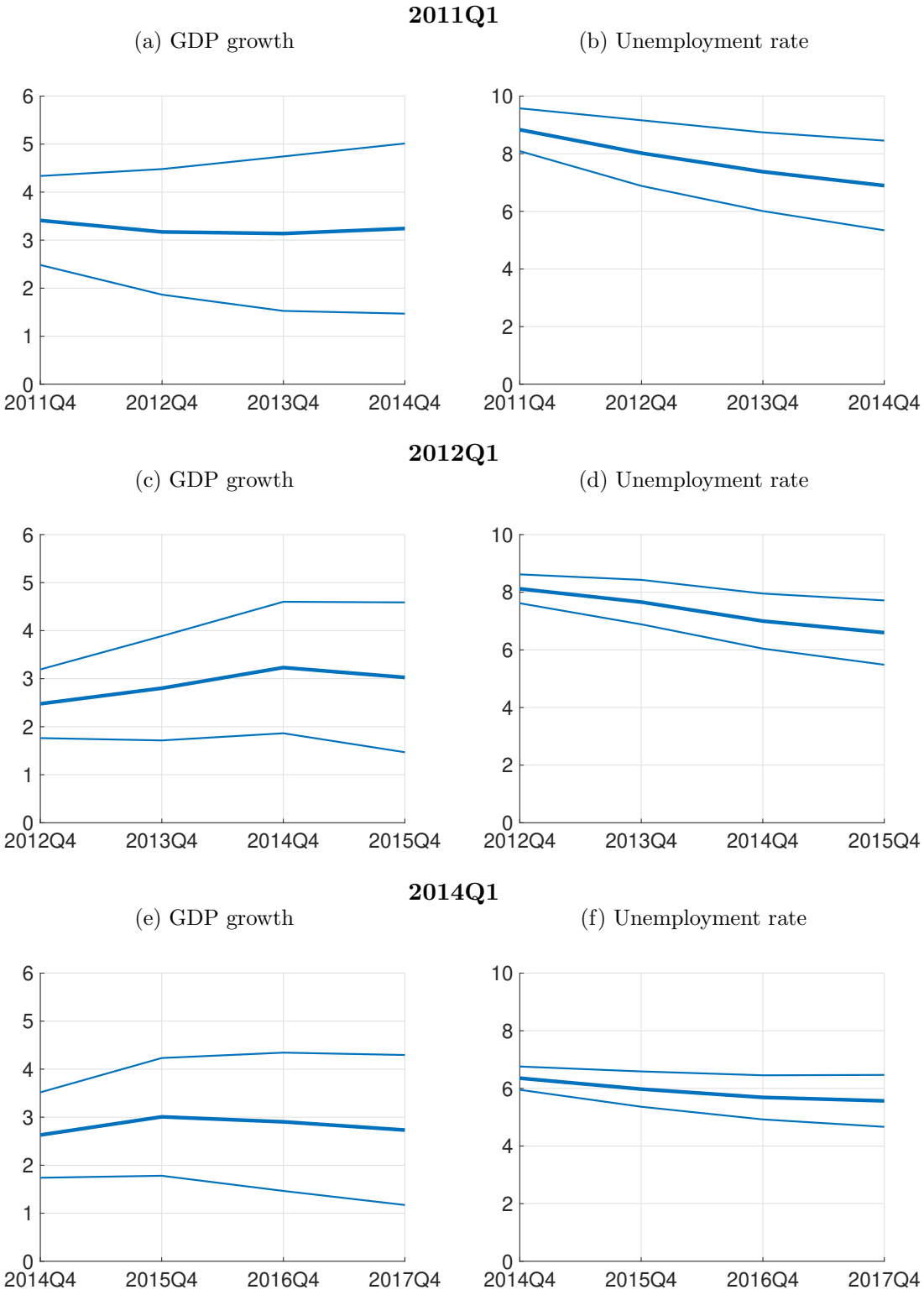
Notes: Cumulative densities obtained from our SV model (with and without entropic tilting), as well as the SPF histograms. The SPF histograms report only discrete probabilities (marked by diamonds), and leave open the precise shape of the underlying density for values in between (as marked by the boxes that are demarcated with blue dotted lines). Probabilities on the y-axis in percentage points, and values for annualized GDP growth in percent on the x-axis.

Figure 8: DRPS of models and SPF, for annual forecasts of GDP growth



Notes: DRPS scores computed for the SPF histograms, as well as SV and CONST models for calendar-year predictions (using the bins defined by the SPF). Left-hand column panels report the time t contribution to the score at each quarter whereas right-hand column panels report average scores computed over growing samples that start in 1992Q1.

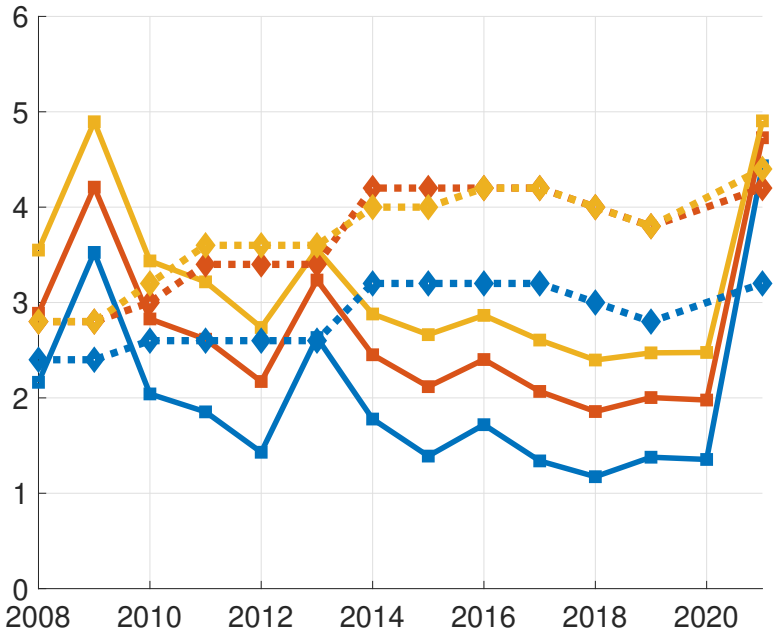
Figure 9: Fan charts for annual predictions from baseline SV model



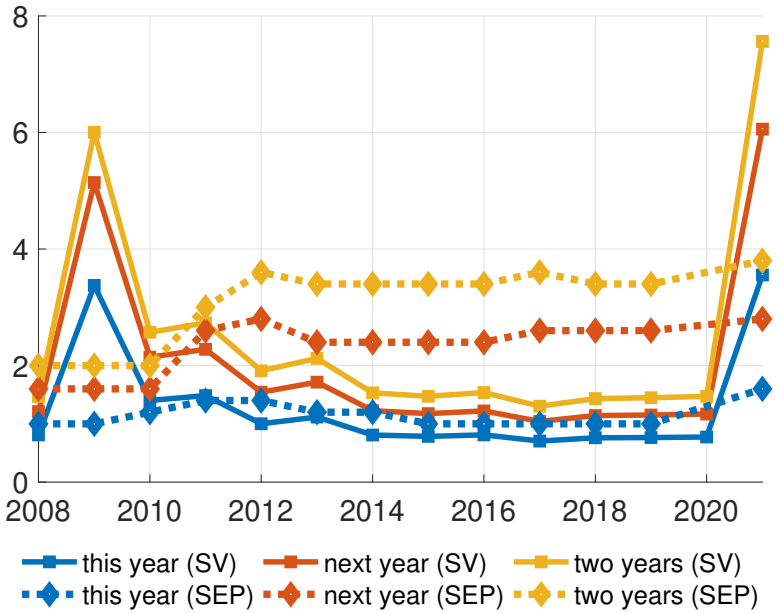
Notes: Predictive means and 68% uncertainty bands generated by the SV model at various forecast origins. For GDP growth, forecasts are for the average growth rate over the targeted and three previous quarters, whereas for the unemployment rate the forecast targets the quarterly level.

Figure 10: Uncertainty measures from SEP and model, annual forecasts

(a) GDP growth



(b) Unemployment rate



Notes: Uncertainty measures as reported by the FOMC's SEP (dash-dotted lines) and generated from our SV model (solid lines). The SEP measure corresponds to two times the RMSE of historical errors of professional forecasters estimated as in (50), and reported in the SEP. The model-based measure is the width of the 68% uncertainty band at the given forecast horizon.

Constructing the Term Structure of Uncertainty 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] *Central Bank of Hungary,* [3] *Deutsche Bundesbank.*

July 8, 2022

Abstract

This online appendix provides additional results to supplement the material in our paper.

List of Tables

S.1 Density forecast comparison: Models vs SPF histograms S.1

List of Figures

S.1 Term structures of unemployment rate expectations S.2

S.2 Forecast fan charts for unemployment rate S.3

S.3 Forecast fan charts for unemployment rate since 2020 . . . . . S.4

S.4 Comparison of GDP growth uncertainty before and after entropic tilting S.5

S.5 Comparison of unemployment rate uncertainty before and after entropic tilting . . . S.6

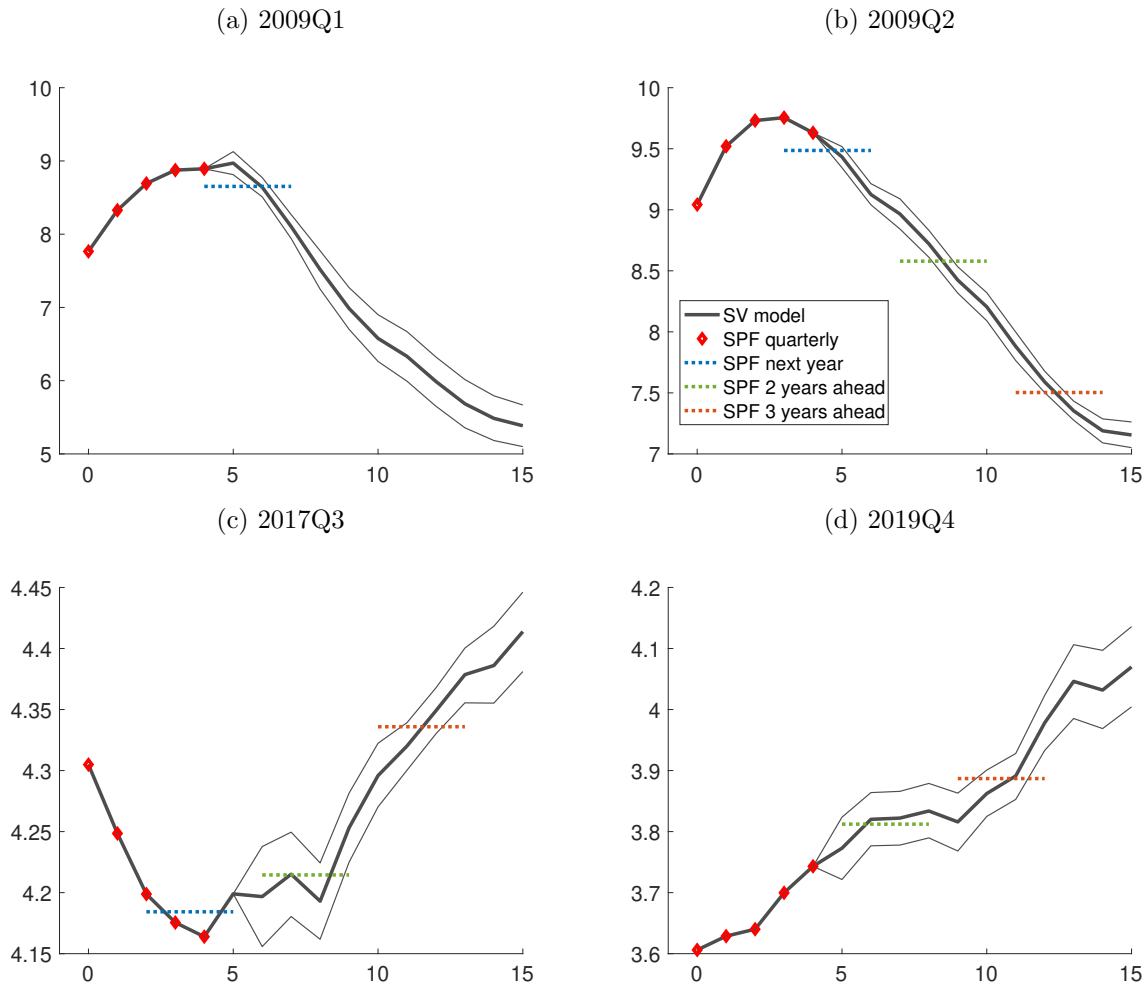
*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland, the Federal Reserve System, the Deutsche Bundesbank, the Central Bank of Hungary, or the Eurosystem.

Table S.1: Density forecast comparison: Models vs SPF histograms

PANEL A: Next year				
	Samples			
Model	1992-2020	1992-2016	2009-2020	2009-2016
SV	1.01	1.00	1.06	1.13**
CONST	1.01	1.00	1.07**	1.17***
PANEL B: Two years ahead				
			Samples	
Model			2009-2019	2009-2016
SV	–	–	1.10***	1.27***
CONST	–	–	1.08*	1.33***
PANEL C: Three years ahead				
			Samples	
Model			2009-2018	2009-2016
SV	–	–	1.08**	1.27***
CONST	–	–	1.08	1.37***

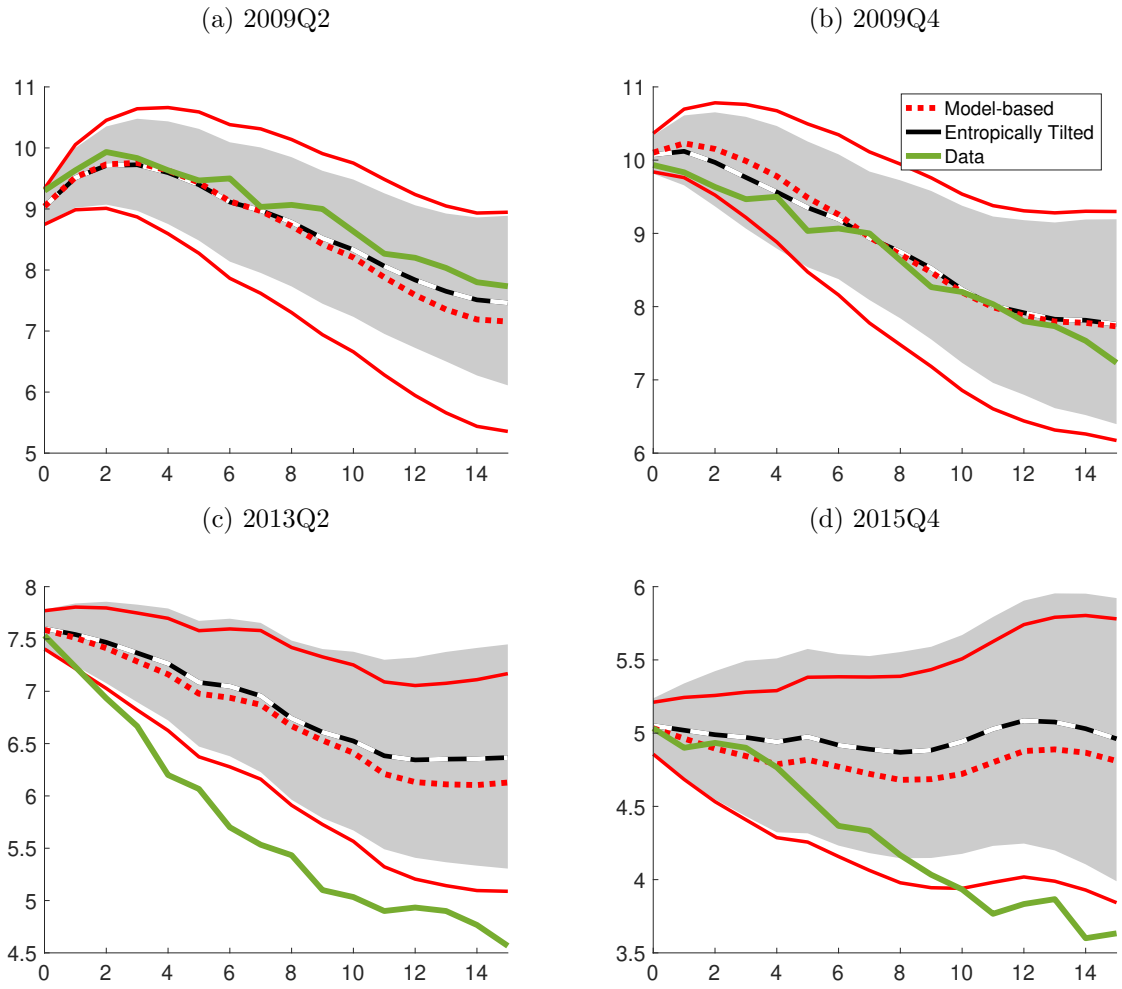
Note: DRPS of model-based calendar-year forecasts for RGDP relative to SPF histograms (and using SPF bins) with SPF in denominator. Quarterly forecast observations. Samples listed at top of each panel refer to first and last year of observations with available SPF forecast and subsequently realized value. Significance assessed by Diebold-Mariano test using Newey-West standard errors with $y \cdot 4$ quarterly lags, where y denotes the value of the annual forecast horizons. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Figure S.1: Term structures of unemployment rate expectations



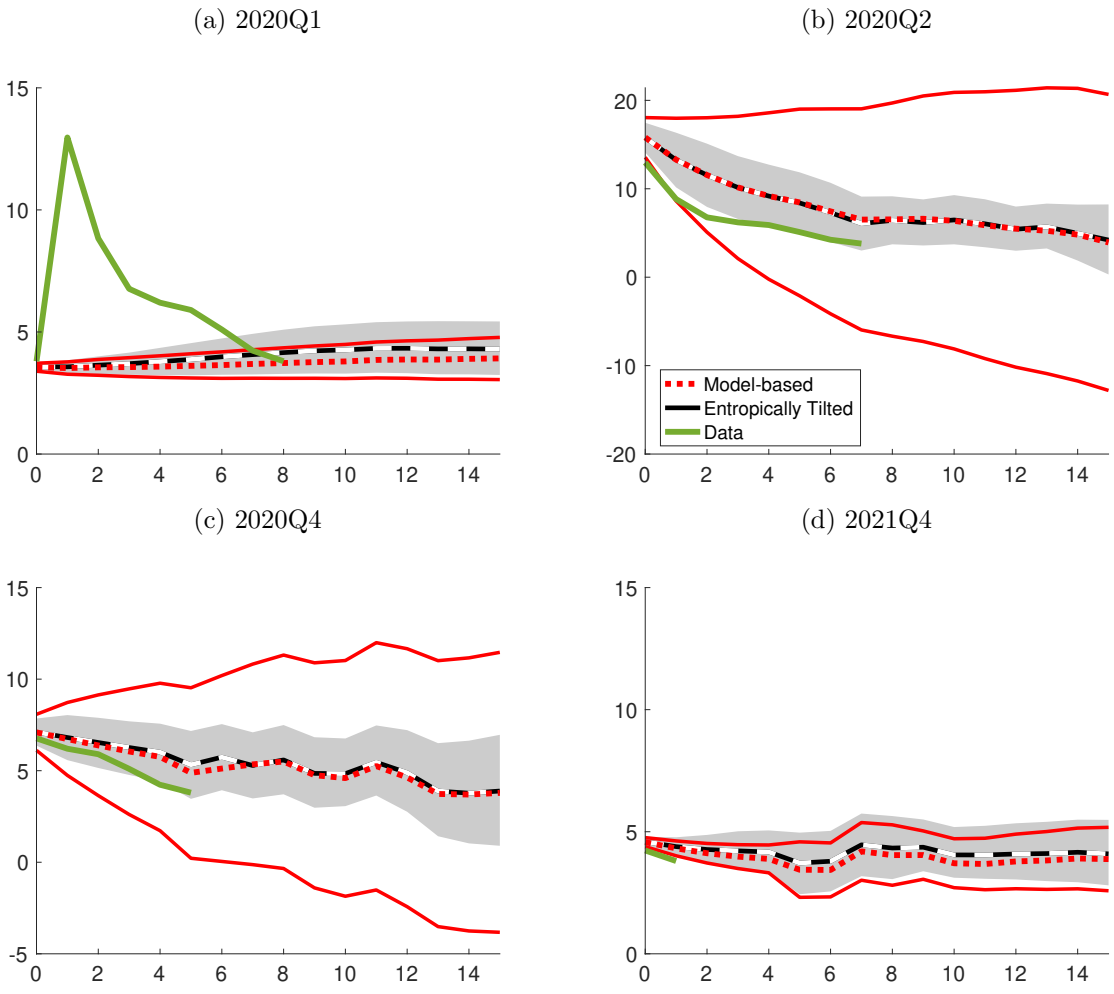
Notes: Term structure of unemployment rate expectations, selected forecast origins. Real-time estimates. Red diamonds give the available quarterly fixed-horizon forecasts from SPF, and dotted lines indicate the available annual fixed-event forecasts (the length of each line corresponds to the quarters included in the tent-shaped linear mapping from quarterly to annual-average changes). Horizons (in quarters) are indicated on the horizontal axis.

Figure S.2: Forecast fan charts for unemployment rate



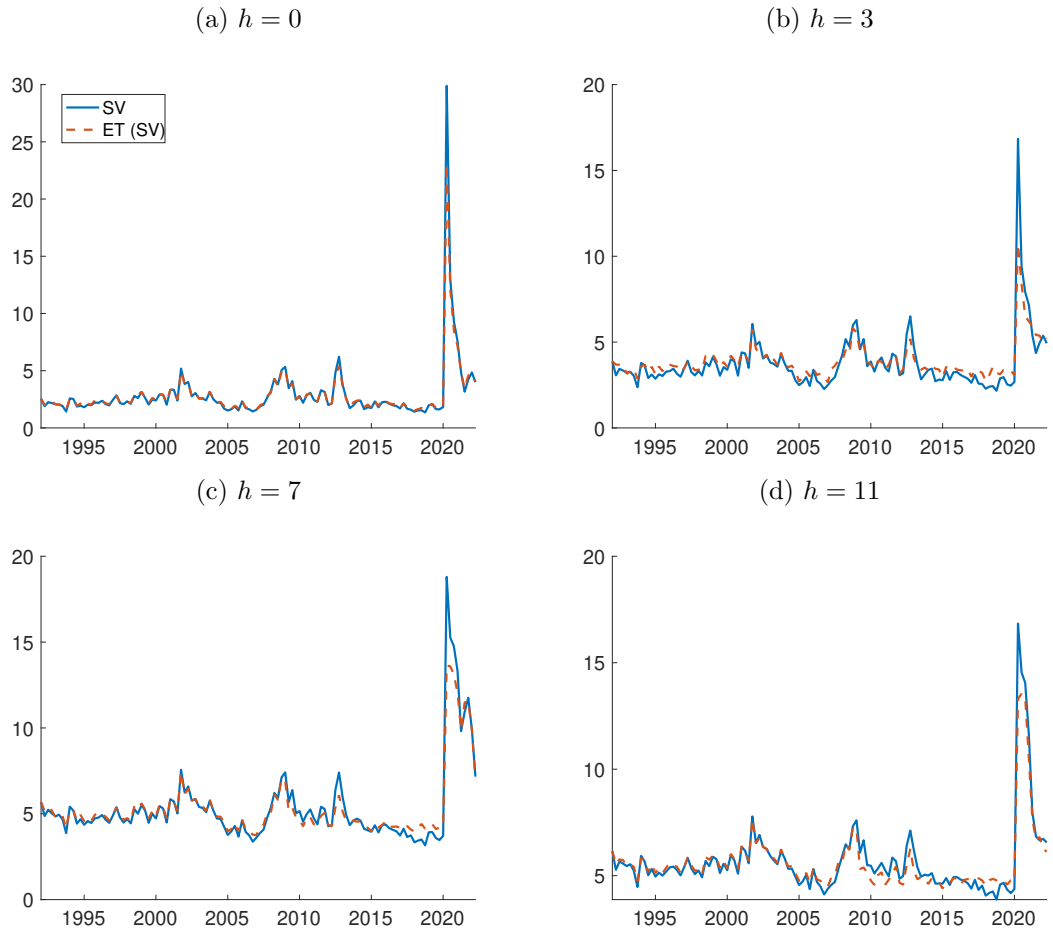
Notes: Forecast fan charts for unemployment rate from the baseline SV model. Model-based (red), entropically tilted (black and grey), with 68% uncertainty bands each, and realized values (green).

Figure S.3: Forecast fan charts for unemployment rate since 2020



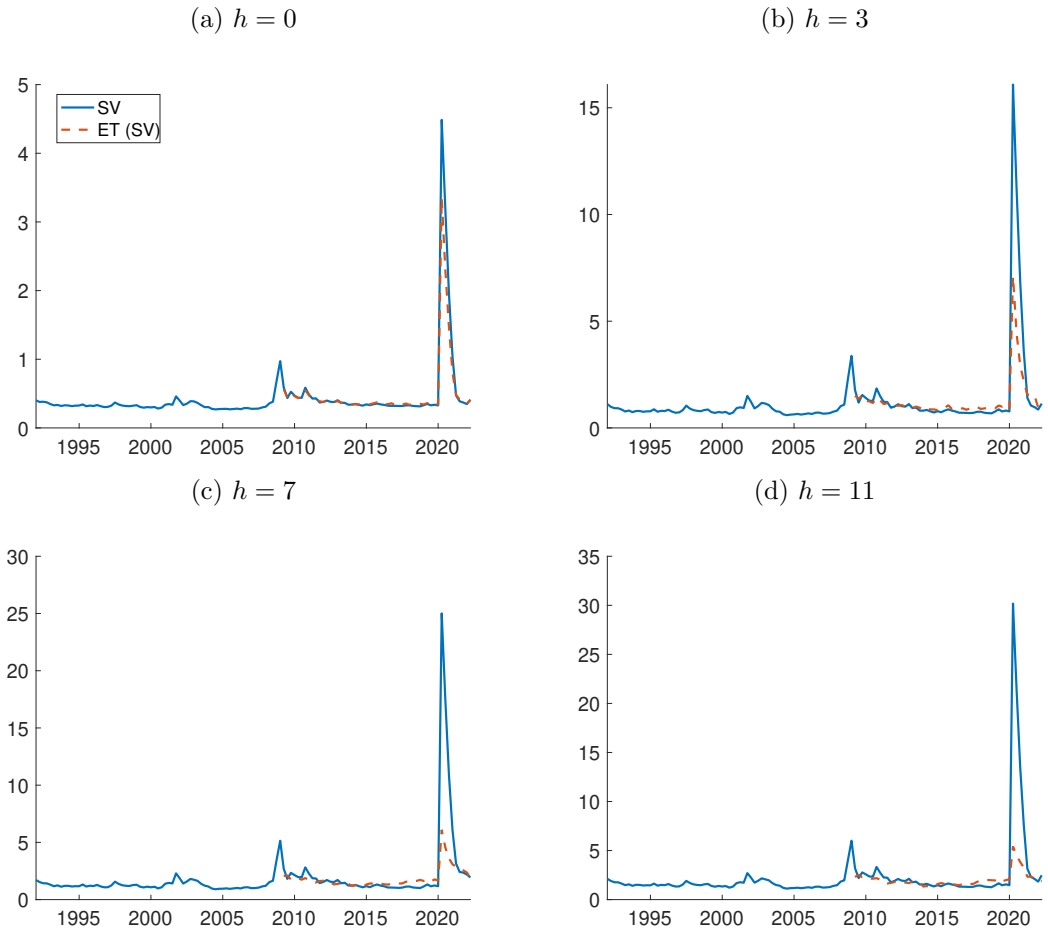
Notes: Forecast fan charts for unemployment rate since 2020 from the baseline SV model. Model-based (red), entropically tilted (black and grey), with 68% uncertainty bands each, and realized values (green).

Figure S.4: Comparison of GDP growth uncertainty before and after entropic tilting



Notes: Comparison of GDP growth uncertainty, measured by the width of 68% bands of predictive real-time densities from the baseline SV model, before and after entroopic tilting to the SPF histograms.

Figure S.5: Comparison of unemployment rate uncertainty before and after entropic tilting



Notes: Comparison of unemployment rate uncertainty, measured by the width of 68% bands of predictive real-time densities from the baseline SV model, before and after entroopic tilting to the SPF histograms.