Federal Reserve Bank of Cleveland Working Paper Series

What is the Predictive Value of SPF Point and Density Forecasts?

Todd E. Clark, Gergely Ganics, and Elmar Mertens

Working Paper No. 22-37

November 2022

Suggested citation: Clark, Todd E., Gergely Ganics, and Elmar Mertens. 2022. "What is the Predictive Value of SPF Point and Density Forecasts?" Working Paper No. 22-37. Federal Reserve Bank of Cleveland. https://doi.org/10.26509/frbc-wp-202237.

Federal Reserve Bank of Cleveland Working Paper Series
ISSN: 2573-7953

What Is the Predictive Value of SPF Point and Density Forecasts?*

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

[1]*Federal Reserve Bank of Cleveland,* [2]*Banco de España and John von Neumann University,* [3]*Deutsche Bundesbank.*

November 21, 2022

Abstract

This paper presents a new approach to combining the information in point and density forecasts from the Survey of Professional Forecasters (SPF) and assesses the incremental value of the density forecasts. Our starting point is a model, developed in companion work, that constructs quarterly term structures of expectations and uncertainty from SPF point forecasts for quarterly fixed horizons and annual fixed events. We then employ entropic tilting to bring the density forecast information contained in the SPF's probability bins to bear on the model estimates. In a novel application of entropic tilting, we let the resulting predictive densities exactly replicate the SPF's probability bins. Our empirical analysis of SPF forecasts of GDP growth and inflation shows that tilting to the SPF's probability bins can visibly affect our model-based predictive distributions. Yet in historical evaluations, tilting does not offer consistent benefits to forecast accuracy relative to the model-based densities that are centered on the SPF's point forecasts and reflect the historical behavior of SPF forecast errors. That said, there can be periods in which tilting to the bin information helps forecast accuracy.

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

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

1 Introduction

Many studies have examined and used point forecasts from professional forecasts such as the US Survey of Professional Forecasters (SPF). The high predictive value of SPF point forecasts is well documented. While quite a few studies point to some persistence in SPF forecast errors (e.g., Coibion & Gorodnichenko (2015) and Bianchi et al. (2022)), there is consensus that SPF point forecasts are hard to beat in real-time accuracy (e.g., Ang et al. (2007), Croushore (2010), Faust & Wright (2013), and Croushore & Stark (2019)).

A number of studies have used point forecasts from the SPF or similar sources 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 & 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. To capture and assess time variation in uncertainty, Jo & Sekkel (2019) estimate a factor stochastic volatility model using errors in fixed-horizon quarterly forecasts from the SPF.

In addition to point forecasts, the SPF provides density forecasts in the form of fixed-event probability bins. More specifically, the SPF publishes (1) fixed-horizon quarterly point forecasts, at shorter horizons, (2) fixed-event annual point forecasts, covering shorter and longer horizons, and (3) fixed-event annual density forecasts in the form of probability bins. Reflecting the challenges to making use of information in fixed-horizon quarterly forecasts and fixed-event annual forecasts, most studies make use of one but not the other. For example, with measurement based on just annual forecasts, Ganics et al. (2021) develop a density combination-based method for translating fixed-event density forecasts from the SPF to fixed-horizon quarterly forecasts. Similarly, regarding point and density forecasts, many studies use one or the other but not both. An exception is Clements & Galvão (2017), who compare the SPF's ex ante density estimates against the ex post root mean squared errors (RMSEs) of the SPF point forecasts and find that the survey's density forecasts overestimate uncertainty relative to the RMSE.

We present a new approach to combining the information in point and density forecasts from the SPF and assess the incremental predictive value of the density forecasts relative to the survey's point forecasts. More specifically, we combine the information in the SPF's quarterly and annual point forecasts and annual probability bins to estimate a quarterly term structure of forecasts and forecast uncertainty that extends out through the longest horizon possible with SPF information. Throughout we use the average predictions published by the SPF, which can be seen as forecasts obtained from a linear pooling of individual SPF participants' forecasts.

We start from model-based density forecasts centered on the SPF's point forecasts, but not informed by the survey's density forecasts. The model is based on our companion work in Clark et al. (2022), henceforth "CGM," and extends Clark et al. (2020), who only used the directly observable term structure of quarterly SPF forecasts. The CGM model combines these fixed-horizon predictions with the SPF's annual fixed-event point forecasts to construct term structures of expectations and uncertainty. (These fixed-event calendar-year outcomes are also the target of the SPF's probability bins.) The CGM framework casts a decomposition of multi-period forecast errors into a sequence of forecast updates that may be partially unobserved, resulting in a multivariate unobserved components model. The model's density predictions are informed by historical SPF forecast errors, which display strong time variation in volatility and are modeled with stochastic volatility.

We employ entropic tilting to bring the density forecast information contained in the SPF's probability bins for fixed-event forecasts to bear on the term structure of expectations and uncertainty. Entropic tilting can be seen as a non-parametric approach to conditional forecasting, used in studies such as Cogley et al. (2005), Ganics & Odendahl (2021), Krüger et al. (2017), Robertson et al. (2005), and Tallman & Zaman (2020). We use tilting to adjust the model-based predictive distribution so that the tilted distribution's implied probability bins match up with those from the SPF (while minimizing a distance criterion).

Tilting allows us to treat the SPF's point and density forecasts commensurately.

The common approach to incorporating information from SPF densities is to first fit a parametric distribution to the SPF histograms and then to proceed based on the first two (or three) moments of the fitted distribution (e.g., Banbura et al. (2021) and Galvão et al. (2021)). Our tilting approach instead directly targets the bin probability; to the best of our knowledge, we are the first to do so. Instead of applying tilting to information obtained from the SPF histograms, other applications have targeted the historical moments of SPF point forecasts and their errors (Krüger et al. (2017)). In our application, historical SPF forecast errors inform the model-based densities that serve as inputs for entropic tilting. The results document the limited merits of the SPF's density predictions compared to densities estimated from the historical errors in SPF point forecasts.

We present empirical results for SPF forecasts of GDP growth and inflation in the GDP price index — variables for which the SPF sample of density forecasts is longest. First, we examine the efficacy of using entropic tilting to incorporate information in the SPF's probability bins for fixed-event annual forecasts, and illustrate the impacts of bin-tilting on predictive distributions. Second, we study the historical forecast accuracy of SPF density predictions themselves and model forecasts informed by the SPF.

Considering the historical forecast accuracy of the SPF's bin forecasts for the next calendar year and beyond, we find these to be on par with our model predictions, at least over the full sample. Similarly, tilting our model densities to the SPF's probability bins yields only modest effects in formal evaluations of forecast accuracy. (This finding applies to both our preferred approach of tilting directly to the bins and the alternative approach of tilting to moments from densities fit to the SPF probability bins.) The model's point and density forecasts — which are conditioned on the historical performance of SPF point forecasts — appear to be good enough that tilting to the SPF's probability bins does not consistently improve forecast accuracy. That said, there can be periods in which tilting to the bin information helps forecast accuracy — for point and density forecasts — as in the case of GDP growth over the period of 2009-2016. On the other hand, for inflation we find slightly detrimental effects from tilting to the SPF histograms over our sample.

Our paper is related to a long line of work concerned with using the SPF's fixed-event probability bins for more general density forecasting purposes. Many have sought — see Clements & Galvão (2017) and references therein — to compare model-based probability densities (or moments thereof) to the bin forecasts. Echoing our findings, Krüger (2017) reports similar forecast performances between histograms and distributions constructed from survey errors for the ECB's SPF. Recently, Bassetti et al. (2022) develop a Bayesian non-parametric approach to density estimation from the bin probabilities. Grishchenko et al. (2019) embed survey-based measures of predictive means and variances into the measurement equation of an affine term structure model with time-varying volatility. Cakmakli & Demircan (2022) improve nowcasts from a factor model of US GDP by adding measurement equations for mean and variance factors that reflect the cross-sectional average and variance (i.e., disagreement) of individual SPF point forecasts. Our finding that adding SPF histogram information to a standard stochastic volatility specification has no consistent advantages for predictive accuracy suggests that model extensions to incorporate extraneous volatility factors will not be needed to capture information conveyed by the SPF. Relatedly, studies including Clements (2018) and Glas & Hartmann (2022) point to potential shortcomings in the predictive accuracy of SPF density forecasts, for example due to rounding of answers by respondents.

The paper proceeds as follows. Section 2 describes the SPF forecasts and data used. Section 3 presents our model and details the entropic tilting that incorporates information from the SPF's probability bins. Section 4 provides results. Section 5 concludes. Additional details and results are provided in a supplementary online appendix.

2 Data

This section first describes our data set of observed SPF point and density forecasts and realized values and then checks the consistency of SPF point and density forecasts.

2.1 SPF forecasts and realized values

Reflecting the forecasts available, we examine quarterly and annual forecasts from the SPF for real GDP growth (RGDP) and inflation in the GDP price index (PGDP). For simplicity, we use "GDP" and "GDP price index" to refer to output and price series, even though, in real time, the measures are based on GNP and a fixed-weight deflator for some of the sample. In all cases, we form the point forecasts and the fixed-event probability bin forecasts using the average over all 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 et al. (2022), simple linear pooling has worked well in other settings. An extension of our work could also focus on alternatives to the linear pooling of individual histogram bins, which might take into account the differing predictive accuracies of individual forecasters as discussed by Genre et al. (2013) and Diebold et al. (2022).

[Table 1 about here.]

We obtained the SPF forecasts from the Federal Reserve Bank of Philadelphia's website. Our estimation samples start with 1968Q4, and the sample end point is 2022Q2. The availability of point and density forecasts at different horizons has considerably changed over time and Table 1 lists the first available dates for SPF predictions of growth and inflation. At each forecast origin, the available fixed-horizon point forecasts typically span five quarters, from the current quarter through the next four quarters. Since 1981Q3, the SPF has included fixed-event point forecasts for the current and next calendar year. In

2009Q2, the forecast horizon for GDP growth (but not inflation) was extended to include annual forecasts for two additional years, i.e., two and three years ahead.

The availability of forecasts for probability bins has also evolved over time. Although the SPF provides probability bins for GDP growth and inflation in the GDP price index starting in 1981, the early years of data pose what Diebold et al. (1999) refer to as "complications," which are also discussed in SPF documentation. These complications include some shifts in the number of bins and their ranges and changes in forecast periods — including some uncertainty as to the horizons of the annual forecasts covered in the 1985Q1 and 1986Q1 surveys. To avoid possible distortions from these issues, we only use probability bin forecasts starting with 1992Q1. In addition, for reasons discussed further below, we also disregard the annual point and density forecasts for the current year.

Within a year of SPF publications, the effective maximum of implied quarterly horizons varies across quarters. For example, the 2021Q4 SPF included fixed-event annual forecasts of GDP growth for the current year and the next three, so that the last annual forecast extends 12 quarters ahead (the annual forecast reported for 2024 includes 2024Q4, 12 quarters beyond the 2021Q4 forecast origin). In the 2022Q1 SPF, the last annual growth forecast for 2025 includes 2025Q4, 15 quarters beyond the forecast origin.

For real GDP and its price index, the SPF solicits point forecasts in levels, whereas density forecasts are surveyed in growth rates. Specifically, point forecasts pertain to quarterly or annual-average levels, which we convert to growth rates based on information included in the survey. For quarterly forecast targets, we use the lagged quarterly level as the basis. To obtain the next-year forecast of annual-average growth, we use the SPF's predictions for the current year as base values (and analogously for the forecasts two and three years ahead). For estimation of the CGM model, growth rates are log-linearized as detailed in Section 3.1 and forecast data are transformed into log differences as well. For our application of entropic tilting, we construct simple growth rates from model output to match the conventions of the SPF histograms, as described further in Section 3.2.

To estimate the CGM model, we also need measures of the outcomes of the variables. From quarterly data files in the Philadelphia Fed's Real-Time Data Set for Macroeconomists (RTDSM), we obtain real-time measures for quarter $t-1$ data that were publicly available to SPF respondents in the quarter t survey. Data on GDP growth and GDP inflation can be substantially revised over time. For forecast evaluation, we measure the outcomes of GDP growth and inflation with the RTDSM vintage published two quarters after the outcome date (that is, we use the quarterly vintage in $t+h+2$ to evaluate forecasts for $t+h$ made in t; this is the second estimate available in the RTDSM's vintages).

2.2 Consistency between SPF point and density forecasts

One aspect of the data that will bear on the impacts of incorporating information in the SPF's probability bin forecasts through entropic tilting is the extent to which the SPF point and density forecasts are mutually consistent. Analyses by Clements (2010, 2014b), Clements et al. (2022), and others (see, e.g., studies referenced by these papers) have documented that, among individual forecasters, inconsistencies between their point forecasts and histograms are common. Clements (2010) emphasizes evidence that forecasters are slower to adjust their probability forecasts than their point forecasts in response to new information. In our case, we are using aggregate rather than individual forecasts, and inconsistencies could be smaller in the former than has been documented for the latter.

[Figure 1 about here.]

Figure 1 compares the annual point forecasts from the SPF to the ranges of possible mean values consistent with the probabilities of the annual bin forecasts. The implied means use the bin probabilities and the bottom, top, and mid-point of each bin to compute lower and upper bounds of histogram-consistent mean forecasts, along with a central tendency. For details, see the supplementary online appendix. These results indicate that, in the aggregate, SPF point forecasts and the means implied by the probability bins are broadly consistent. For 1-year-ahead forecasts of GDP growth, the reported point forecasts are generally comparable to the central tendency computed with the mid-points of the SPF bins, sometimes a little higher and sometimes a little lower. Similarly, the 1-year-ahead forecasts of inflation appear consistent, although from 1992 until about 2012, the point forecasts often run below the central tendency values implied by the probability bins.

For annual forecasts of GDP growth at longer horizons, the SPF's point forecasts consistently lay above the mid-points of the SPF's probability bin ranges but generally remain between the top and bottom of the ranges for mean values consistent with the histograms. Given these differences, it may be the case that entropic tilting to the SPF's bin forecasts will have more impact on longer horizon growth forecasts than shorter horizon forecasts. In particular, in these cases, tilting to the bins may pull down the means of forecast distributions compared to the entirely model-based means. But for inflation, in the early years of the sample, tilting to the bins may pull up forecast means compared to the entirely model-based predictions.

3 Time Series Model and Entropic Tilting Method

As input to our application of entropic tilting, we use a multivariate unobserved components model developed in CGM to generate term structures of expectations and uncertainty that are centered on SPF point forecasts (but not informed by SPF density forecasts). After a brief overview of the model, this section describes the entropic tilting methods used in our paper to align the model-based densities with the SPF's density predictions.

3.1 A model for survey expectations and uncertainty

We apply entropic tilting to output from the CGM state space model that is specified and estimated on a variable-by-variable basis. The variables considered in our paper are real GDP and the GDP price index. For each, we denote their quarterly growth rate (measured as annualized change in log levels) by y_t, and a forecast made at t for quarter $t+h$ is $y_{t+h|t}$. The maximum quarterly horizon that can be covered in the historical SPF data is H. CGM model the evolution of a partially latent state vector $\boldsymbol{Y_t} \equiv \left(y_{t-1}, y_{t|t}, y_{t+1|t}, \ldots, y_{t+H|t}\right)'$ that consists of the lagged realized value y_{t-1} (as observed at t) and the time t term structure of expectations for horizons $h = 0$ through H. In its baseline version, the state space model has the following form:

$$\boldsymbol{Z_t} = \boldsymbol{C_t}\,\boldsymbol{Y_t}\,, \qquad \boldsymbol{Y_t} = \boldsymbol{F}\;\boldsymbol{Y_{t-1}} + \boldsymbol{\eta}_t\,, \qquad \boldsymbol{\eta}_t \sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_t\right), \tag{1}$$

where the measurement vector $\boldsymbol{Z_t}$ collects SPF point forecasts observed at t, and the matrices $\boldsymbol{F}$ and $\boldsymbol{C_t}$ are known based on data definitions. As described next, $\boldsymbol{\eta}_t$ is a vector of forecast updates. We follow the baseline version of the CGM model and treat the vector $\boldsymbol{\eta}_t$ as a martingale difference sequence, $E_{t-1}\boldsymbol{\eta}_t = \boldsymbol{0}$, with a stochastic volatility specification for its variance matrix $\boldsymbol{\Sigma}_t$.

The model builds on an accounting identity that decomposes h-step-ahead forecast errors into the sum of the $t+h$ nowcast error and preceding forecast updates:

$$y_{t+h} - y_{t+h|t} = e_{t+h} + \sum_{i=1}^{h} \mu_{t+h|t+i}\,, \tag{2}$$

with $e_{t+h} \equiv y_{t+h} - y_{t+h|t+h}$ and $\mu_{t+h|t+i} \equiv y_{t+h|t+i} - y_{t+h|t+i-1}$, so that e_{t+h} is the nowcast error at $t+h$, and $\mu_{t+h|t+i}$ measures the update in the forecast of y_{t+h} at time $t+i$. We denote changes to the longest-horizon forecast, $y_{t+H|t}$, as μ_t^*:

$$\mu_t^* \equiv y_{t+H|t} - y_{t+H-1|t-1}\,. \tag{3}$$

After collecting e_{t-1}, $\left\{\mu_{t+h|t}\right\}_{h=0}^{H}$, and μ_t^* in the vector $\boldsymbol{\eta}_t$, we obtain the transition equation in (1) with $\boldsymbol{F}$ being a known matrix of zeros and ones.

Based on the strong evidence for time-varying volatility demonstrated by CGM, we adopt their baseline specification (denoted SV) of stochastic volatility in the process for $\boldsymbol{\eta}_t$. The model decomposes forecast updates into long-run shifts and cyclical gaps. Changes in the long-run forecast, μ_t^*, are assumed to have constant variance, while a scalar stochastic volatility process affects the cyclical gaps at all horizons other than H:

$$\boldsymbol{\eta}_t = \begin{bmatrix} \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} \\ 0 \end{bmatrix} + \mathbf{1} \cdot \boldsymbol{\mu}_{\boldsymbol{t}}^*, \qquad \boldsymbol{\mu}_{\boldsymbol{t}}^* \sim N(0, \sigma_*^2), \qquad \tilde{\boldsymbol{\eta}}_{\boldsymbol{t}} \sim N\left(\mathbf{0}, \lambda_t \cdot \tilde{\boldsymbol{\Sigma}}\right), \tag{4}$$

$$\log \lambda_t = \delta \log \lambda_{t-1} + \nu_t, \qquad \nu_t \sim N(0, \sigma_\nu^2),$$

where $\tilde{\boldsymbol{\eta}}_{\boldsymbol{t}}$ is a vector of forecast-update gaps with $N_y - 1$ elements. The log SV process has a mean of 0 and slope coefficient δ to be estimated. The time-varying variance with mean of 1 scales up a full variance-covariance matrix $\tilde{\boldsymbol{\Sigma}}$. Although the ordering of variables commonly affects estimates of VARs with SV processes for each variable (see discussions in studies such as Arias et al. (2022)), with the common SV specification, the ordering of variables in the model has no impact on estimates. Additionally, our empirical analysis includes comparisons to estimates from a model labeled as CONST, which treats the innovation vector $\boldsymbol{\eta}_t$ as conditionally homoskedastic, with variance-covariance matrix Σ.

For growth in real GDP and the GDP price index, CGM employ the following log-linear approximation to model annual-average changes:

$$\hat{y}_t = 1/16 \cdot (y_t + 2 \cdot y_{t-1} + 3 \cdot y_{t-2} + 4 \cdot y_{t-3} + 3 \cdot y_{t-4} + 2 \cdot y_{t-5} + y_{t-6}) \ , \tag{5}$$

where $\hat{y}_t$ denotes the (log-linearized) growth rate of the annual-average levels of the year ending in quarter t over the year ending in $t-4$. When t corresponds to a Q4 observation, $\hat{y}_t$ captures a calendar-year change. Other examples relying on such approximations include Aruoba (2020), Mariano & Murasawa (2003), and Patton & Timmermann (2012).

The variables $y_{t+h|t}$ and $\hat{y}_{t+h|t}$ denote survey expectations collected at forecast origin t for forecast targets y_{t+h} and $\hat{y}_{t+h}$, respectively. The measurement vector $\boldsymbol{Z}_{\boldsymbol{t}}$ contains the available SPF forecasts for fixed horizons ($y_{t+h|t}$) and fixed events ($\hat{y}_{t+h|t}$), as well as a real-time reading of the last realized value, y_{t-1}. Since the SPF provides fixed-horizon forecasts for up to four quarters ahead, current-year (fixed-event) forecasts are disregarded by CGM. In a similar vein, next-year forecasts are ignored when published in the fourth quarter. Otherwise, $\boldsymbol{Z_t}$ includes all available readings of fixed-event forecasts for the next year and beyond. Collecting terms, the measurement equation has the form given in (1), with the elements of $\boldsymbol{C_t}$ known and reflecting data definitions and transformations discussed above, as well as shifts in data availability.

As detailed by CGM, the model is estimated with a Gibbs sampler using joint data for observed realizations and SPF predictions for a given economic variable. We retain 3,000 draws after a burn-in 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 draw, resulting in $S = 300,000$ predictive density draws.

3.2 Entropic tilting

As noted in the introduction, entropic tilting has gained popularity as a convenient post-estimation method to incorporate additional moment conditions into a model's predictive distribution. 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. 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 (KL) sense.

Let $\boldsymbol{X}_t = \left[x_t^1, \ldots, x_t^n, \ldots, x_t^N\right]'$ collect predictions for the rates of change in calendar-year average levels of RGDP (or PGDP) at forecast origin t for calendar years $1, \ldots, N$ ahead. For example, superscript 1 corresponds to the next calendar year.

To construct predictive densities for calendar-year events, we need to map draws from the predictive density of the quarterly model into annual growth rates. Let $\mathcal{Y}_t$ denote the level of RGDP in quarter t. As before, $y_{t+h} = 400(\log \mathcal{Y}_{t+h} - \log \mathcal{Y}_{t+h-1})$ denotes the annualized quarterly growth rate of RGDP from quarter $t-1$ to t and we let $R_{t+h} \equiv \exp\left(y_{t+h}/400\right) = \mathcal{Y}_{t+h}/\mathcal{Y}_{t+h-1}$ denote the corresponding quarterly gross rate of change. For simplicity, consider calculating the growth rate of average annual RGDP 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 rate can be expressed as

$$\begin{aligned} x_t^1 &= 100 \cdot \left(\frac{\mathcal{Y}_{\tau+5} + \mathcal{Y}_{\tau+6} + \mathcal{Y}_{\tau+7} + \mathcal{Y}_{\tau+8}}{\mathcal{Y}_{\tau+1} + \mathcal{Y}_{\tau+2} + \mathcal{Y}_{\tau+3} + \mathcal{Y}_{\tau+4}} - 1 \right) \\ &= 100 \cdot \left(\left(\prod_{j=2}^{5} R_{\tau+j} \right) \cdot \frac{1 + \sum_{k=2}^{4} \prod_{j=1}^{k} R_{\tau+4+j}}{1 + \sum_{k=2}^{4} \prod_{j=1}^{k} R_{\tau+j}} - 1 \right) . \end{aligned} \tag{6}$$

Based on equation (6), 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 were 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 annual growth rates featured in the SPF is thus free of log-linear approximations.

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 Markov chain Monte Carlo (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 $\widetilde{f}_t$. 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 divergence, to the original model-based distribution f_t. Formally, entropic tilting is a minimization with KL as the objective function subject to constraints that represent the moment conditions derived from the SPF bins (which f_t does not satisfy in general).

Let $g : \mathbb{R}^N \rightarrow \mathbb{R}^{\dim(\bar{\boldsymbol{g}}_t)}$ denote the function that maps the MCMC draws into a vector whose expected value under the tilted distribution $\widetilde{f}_t$ we want to set equal to $\bar{\boldsymbol{g}}_t$. Entropic tilting ensures that under the tilted distribution $\widetilde{f}_t$, we have $E_{\widetilde{f}_t} g\left(\boldsymbol{X}_t\right) = \bar{\boldsymbol{g}}_t$, and $E_{\widetilde{f}_t} g\left(\boldsymbol{X}_t\right) \equiv \sum_{s=1}^{S} \widetilde{w}_s \cdot g\left(\boldsymbol{X}_t^s\right)$ denotes the expected value of $g\left(\boldsymbol{X}_t\right)$ under the distribution $\widetilde{f}_t$. Entropic tilting is then the solution of the following optimization problem:

$$\min_{\{\widetilde{w}_s\}} \mathrm{KL}(\widetilde{f}_t, f_t) = \sum_{s=1}^{S} \widetilde{w}_s \cdot \log\left(\frac{\widetilde{w}_s}{w_s}\right) \qquad \text{such that} \qquad E_{\widetilde{f}_t}\, g(\boldsymbol{X}_t) = \bar{\boldsymbol{g}}_t\,, \tag{7}$$

where the weights $\{\widetilde{w}_s\}_{s=1}^{S}$ characterizing $\tilde{f}_t$ need to be non-negative and sum to one. The minimizing solution is

$$\widetilde{w}_s^* = \frac{\exp\left(\gamma^{*\prime} g\left(\boldsymbol{X}_t^s\right)\right)}{\sum_{s=1}^{S} \exp\left(\gamma^{*\prime} g\left(\boldsymbol{X}_t^s\right)\right)}\,, \quad \text{with} \quad \gamma^* = \underset{\gamma}{\operatorname{argmin}} \sum_{s=1}^{S} \exp\left(\gamma'\left(g(\boldsymbol{X}_t^s) - \bar{\boldsymbol{g}}_t\right)\right)\,. \tag{8}$$

The SPF probability forecasts are available as probabilities assigned to pre-specified bins. For example, in the 2013Q1 SPF round, panelists were asked to provide the probability of annual real GDP growth falling into the following 11 bins: < -3, -3 to -2.1, -2 to -1.1, -1 to -0.1, 0 to 0.9, 1 to 1.9, 2 to 2.9, 3 to 3.9, 4 to 4.9, 5 to 5.9, and ≥ 6, all in average annual percentage points between calendar years 2013 and 2016. 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 the MCMC draws from the 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 $b_2 = [-3.0, -2.05)$ and $b_3 = [-2.05, -1.05)$, respectively. Second, while SPF panelists also submit predictions for the current calendar year, we do not utilize this information. Studies including Clements (2014a), Clements (2018) , and Clements & Galvão (2017) have shown that the SPF's probability bins overstate forecast uncertainty at shorter horizons, in the sense that ex ante bin-based measures exceed measures based on historical errors in point forecasts from the SPF or time series models. This evidence suggests a particularly poor predictive value of the SPF's current-year densities, which is confirmed by results in the supplementary online appendix. While we find SPF density predictions for next year and beyond to be quite competitive compared to our model, the probabilistic forecasts conveyed by the current-year bins strongly fail in comparison to our model densities trained on historical forecast errors.

Much of the literature fits parametric distributions to the SPF histograms or makes semi-parametric assumptions to use them. Differently from that approach, as a novelty of our paper, we directly use the information contained in the SPF histograms in our baseline exercise, without making any distributional assumptions. We see this as an advantage of our approach, adhering to the information contained in the SPF. 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 ECB's documentation of the Bayesian Estimation, Analysis, and Regression toolbox (Dieppe et al. (2016)).

3.2.1 Tilting to histograms

Probabilities can be written as expectations over indicator functions, and thus as moments. Specifically, for event $\mathcal{A}$, Prob $(\mathcal{A}) = E\left(\mathbb{1}\left(\mathcal{A}\right)\right)$, where $\mathbb{1}\left(\cdot\right)$ is the indicator function. Let $g \equiv [p_1, \ldots, p_n, \ldots, p_N]'$, where $p_n(x_t^n) = \left[\mathbb{1}\left(x_t^n \in b_1\right), \ldots, \mathbb{1}\left(x_t^n \in b_B\right)\right]'$, and we can impose the probabilities in the SPF bin forecasts as moments on the tilted distribution, by collecting the SPF bin probabilities in the vector $\bar{\boldsymbol{g}}_t$. The dimension of $\bar{\boldsymbol{g}}_t$ is the product of (a) the number of calendar years for which predictions are used, and (b) the number of bins (less one, since probabilities sum to one for a given calendar-year target). Staying with the example, in 2013Q1, we use calendar-year forecasts for $N = 3$ years, with $B = 10$ bins each, leading to $N \cdot B = 30$ moment conditions.

3.2.2 Tilting to fitted moments

Several previous papers utilized information in survey histograms by fitting a parametric distribution and tilting a model's predictive density to the moments of the fitted distribution (see, e.g., Banbura et al. (2021) and Galvão et al. (2021)). As a robustness check of our baseline approach that applies tilting directly to the SPF bin probabilities and does not require additional distributional assumptions, we also consider tilting to higher-order moments of a fitted distribution. We follow Engelberg et al. (2009) and more recent studies such as Galvão et al. (2021) and Krüger & Pavlova (2022) in fitting a generalized beta distribution to the cumulative histogram via non-linear least squares at each forecast origin t and forecast horizon n.

Based on the estimated parameters of the generalized beta distributions' fit at each forecast origin, we calculated the mean m_t^n, variance v_t^n, and central skewness sk_t^n of the distribution. To apply entropic tilting to these moment conditions, we construct the following $3 \cdot N$ vectors, where N denotes the number of annual forecasts used:

$$g = [g_{1,1}, \ldots, g_{N,1}, g_{1,2}, \ldots, g_{N,2}, g_{1,3}, \ldots, g_{N,3}]' \,, \tag{9}$$

$$\text{and} \quad \bar{\boldsymbol{g}}_t = \left[m_t^1, \ldots, m_t^N, v_t^1 \ldots, v_t^N, \text{sk}_t^1 \ldots, \text{sk}_t^N\right]', \tag{10}$$

with $g_{n,1} = x_t^n$, $g_{n,2}(x_t^n) = (x_t^n - m_t^n)^2$, and $g_{n,3}(x_t^n) = ((x_t^n - m_t^n)/\sqrt{v_t^n})^3$. In the interest of brevity, the supplementary online appendix further details the fitting procedures and tilting implementation and the robustness of our results to the selection of targeted moments and to fitting the bins to a normal distribution instead of the generalized beta.

4 Results

Incorporating information from the SPF's annual fixed-event probability bins will affect the predictive densities from the 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 term structure of forecast uncertainty and the historical accuracy of SPF forecasts, at both annual and quarterly horizons. For brevity, this section provides complete results for real GDP growth and more selected results for inflation in the GDP price index. The supplementary online appendix reports selected results for the unemployment rate.

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

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 GDP growth using both 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.

[Figure 2 about here.]

Figure 2 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 the 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.

[Figure 3 about here.]

In another example, Figure 3 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.

4.2 Impacts of bin tilting on the term structure of uncertainty

To more directly assess changes in uncertainty across horizons and over time and impacts of tilting to the SPF's probability bins, Figure 4 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, panel (a) includes a subset of quarterly horizons. Other panels compare the model-based estimates to those that incorporate entropic tilting to the SPF's probability bins.

[Figure 4 about here.]

The model-based estimates in panel (a) of Figure 4 show two general patterns. First, as expected, forecast uncertainty tends to rise with the forecast horizon. Uncertainty is noticeably higher at longer horizons (7 or more quarters) than shorter horizons (0 to 3 quarters). From 0 to 3 quarters, uncertainty gradually increases. From 7 to 15 quarters, uncertainty continues to rise, but typically by less than in the short horizon case. Second, the uncertainty of out-of-sample forecasts of GDP growth fluctuates significantly over time. After 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.

The comparisons of model-based estimates to their tilted counterparts in panels (b) through (d) of Figure 4 show that the same general patterns apply to estimates of forecast uncertainty informed by tilting to the SPF's probability bins. For most of the period, the tilted estimates that incorporate information from the SPF's probability bins are very similar to the entirely model-based estimates that rest on just SPF point forecasts. The bin information can push uncertainty a little above the model estimates in some periods and below in others. The largest impact of bin tilting occurs in the early period of the COVID-19 pandemic, when the information from the bins (which can reflect the subjective judgment of the average survey respondent) helps mitigate the rise in uncertainty that occurred as macroeconomic volatility temporarily soared.

4.3 Entropic tilting's impacts on historical forecast accuracy

Ultimately, we are interested in whether the forecasts we construct using the CGM 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 sometimes small. In this section we turn to a more formal assessment of the broader question at hand. For these results, from 1992Q1 onward, out-of-sample forecasts are generated for all quarterly horizons from $h = 0$ to 15 for GDP growth and from $h = 0$ to 7 for inflation, 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 by transforming them to simple as opposed to logarithmic growth rates, in line with the SPF definition. This comparison relies on the (discrete) rank probability score (DRPS), which has been been applied in various studies of survey forecasts, including Boero et al. (2011), Clements (2018), and Krüger & Pavlova (2022). The DRPS assigns scores based on outcomes being within bins or not, with $\text{DRPS}_t = \sum_{k=1}^{K} \left(P_t^k - D_t^k\right)^2$, where K denotes the number of probability bins, P_t^k is the cumulative bin forecast probability from bins 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 1-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 1-year-ahead horizon. This likely reflects the pattern noted earlier that, over the period, the mean forecasts implied by the central tendencies of the SPF bins were lower than the SPF's point forecasts, which improves accuracy over a period in which growth outcomes were relatively low by historical standards.

[Figure 5 about here.]

To shed more light on the DRPS accuracy of the SPF bins as compared to the models, Figure 5 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 1-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. Mechanically, over this period, at longer annual horizons the mean forecasts implied by the probability bins tended to lay below the SPF's point forecasts. With the economy growing more slowly in the early years of the recovery from the Great Recession than projected in the point forecasts, the tilting to the bins that imply lower mean growth helps to improve the accuracy of the forecasts.

[Table 2 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 CGM model, using RMSE for point forecasts and the continuous ranked probability score (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. Statistical significance is assessed using the Diebold & Mariano (1995) test with Newey & West (1987) standard errors. The first two columns of the table provide the raw levels of scores from the SV baseline.

Before we take up the efficacy of tilting to the bins, it is worth noting that the CGM baseline model with SV has some advantages over their CONST specification, particularly in density accuracy. 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 are identical to the SPF forecasts). However, the density forecasts of the CONST specification are consistently less accurate than those of the baseline SV model (as measured by CRPS). For example, in the pre-pandemic sample, the accuracy gains of SV are roughly 2 to 3 percent at shorter horizons and 5 to 7 percent at longer horizons, which are often statistically significant. These benefits from SV are consistent with the findings of earlier studies (e.g., Clark (2011) and D'Agostino et al. (2013)) that have found SV to regularly yield improvements in time series forecasts.

As to the efficacy of incorporating bin forecasts through entropic tilting, starting with point forecasts, incorporating the information in the SPF probability bins through entropic tilting has little impact on predictive accuracy. 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. 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. The result is not too surprising, since both SV and CONST model generate forecasts that are centered on the SPF point forecasts, which are consistent with the bins as shown in Section 2.

A more striking result is that the impacts of tilting on density forecast accuracy are slim as well. 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. 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. As a corollary, the tilted outputs from the SV and CONST models remain quite different, suggesting some limits to the information conveyed by the SPF histograms.

Overall, we read the evidence as indicating that the impacts — on forecast accuracy — of incorporating the information in the SPF's probability bins through entropic tilting are modest. We see the inability of entropic tilting to deliver significant gains either in point or density forecasts beyond the 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. As shown in the supplementary online appendix, for the subsample of 2009-2016, tilting to the SPF bins also improves the accuracy of quarterly point and density forecasts of GDP growth (relative to our model). Measured by RMSE and CRPS, the gains to tilting over this short period in the wake of the financial crisis are on the order of 5 percent. It is also possible that our overall evidence reflects inherent challenges in using SPF-like probability bins for annual forecasts to convey rich information on quarterly forecast densities. A robustness check in the supplementary online appendix examines the accuracy of forecasts from the CONST specification tilted to match annual forecast probabilities from the SV model. These tilted CONST forecasts fall well short of the accuracy of the SV forecasts themselves.

[Table 3 about here.]

Table 3 provides corresponding results on the historical accuracy of quarterly forecasts of inflation in the GDP price index, comparing our baseline SV specification to the CONST model and their counterparts entropically tilted to the SPF's probability bins. Because the SPF's reported probability bins for inflation only cover the next year and not subsequent years, the covered horizons end with 7 quarters ahead.

The results for GDP inflation share many qualities with those for GDP growth, but also display marginally negative impacts of tilting. First, the inclusion of stochastic volatility in the model improves the accuracy of density forecasts and has a smaller impact on the accuracy of point forecasts. At the longer horizons covered, the RMSEs of the CONST specification are 3 to 6 percent higher than the RMSEs of the SV baseline. The CRPS ratios for CONST point to larger density accuracy gains of SV, rising from around 4 percent at short horizons to as much as 22 percent at the longest horizon.

Second, entropic tilting to the SPF's probability bins is not only unable to deliver gains in the accuracy of point or density forecasts of inflation, but also delivers some noticeable losses. Over the samples reported in the table, by both the RMSE and the CRPS metrics, the tilted SV forecasts are roughly 3 to 5 percent less accurate than the model's forecasts. This seemingly surprising result could be explained by noting that, in panel (d) of Figure 1, the mid-range implied by SPF bins runs above the point forecast (directly utilized by the model), which, coupled with downside surprises in realized inflation, rewards the model-based forecasts centered on SPF point predictions. Furthermore, these results are in line with those of earlier studies (see, e.g., Mertens (2016)) demonstrating that survey point forecasts were slow to track the fall in realized inflation over the 1990s, and the histograms were lagging even more (see, e.g., Clements (2010)).

4.4 Robustness check: Tilting to fitted moments

As a robustness check of our approach of tilting to histograms, we have also investigated the impact of entropically tilting the model's forecasts to the first three moments of fitted generalized beta distributions (mean, variance, and skewness).

[Table 4 about here.]

Table 4 shows that tilting to fitted moments yields quarterly forecasts that are either no more accurate (real GDP growth) or less accurate (inflation in the GDP price index) than those obtained by tilting directly to the bins. This pattern applies with both the SV and the CONST specifications. In the GDP growth results in the upper panel, the RMSE and CRPS ratios are 1 in most cases, with a few exceptions of 1 percent gains or losses. In turn, our broad finding that the forecast accuracy impacts of incorporating the information in the SPF's probability bins through entropic tilting are modest is robust to the alternative approach of tilting. In the lower panel's results for inflation in the GDP price index, the RMSE and CRPS ratios that in many cases fall between 1.01 and 1.03 indicate that point and density accuracy are slightly lower with the alternative tilting approach than with our baseline approach of tilting directly to the SPF's probability bins. It follows that our baseline finding is robust to the alternative approach of tilting; making use of density forecast information from the SPF fails to deliver gains in the accuracy of point or density forecasts of inflation. (The supplementary online appendix reports similar results from further robustness checks.) Given this robustness and the similar results that obtain with the two tilting approaches, we prefer tilting to histograms instead of tilting to moments derived from a specific distribution, as the former method does not require the econometrician to impose additional parametric assumptions on the SPF bins to obtain the moments.

5 Conclusion

This paper presents a new approach to combining the information in point and density forecasts from the SPF and assesses the incremental predictive value of the density forecasts relative to the survey's point forecasts. We start from model-based density forecasts centered on the SPF's point forecasts, but not informed by the survey's density forecasts. The model, based on our companion work in Clark et al. (2022), combines the SPF's fixed-horizon quarterly and annual fixed-event point forecasts to construct term structures of expectations and uncertainty. We employ entropic tilting to bring the density forecast information contained in the SPF's histograms to bear on the term structure of expectations and uncertainty. As a novelty, we directly match the SPF's bin probabilities with their model-based counterparts and without further parametric assumptions.

Our empirical analysis using SPF forecasts of GDP growth and inflation finds limited merits of the SPF's density predictions compared to densities estimated from the historical errors in SPF point forecasts. Over the full sample, the historical forecast accuracy of the SPF's bin forecasts for the next calendar year and beyond is on par with the model-based forecast accuracy. Similarly, in formal evaluations of quarterly forecast accuracy, tilting the model densities to the SPF's probability bins yields only modest effects. The model's point and density forecasts appear to be good enough that incorporating information from the SPF's probability bins does not consistently improve forecast accuracy. For inflation, incorporating information from the SPF's probability bins actually reduces forecast accuracy. But there can be instances in which tilting to the bin information helps forecast accuracy, as in the case of GDP growth over the period of 2009-2016.

References

Ang, A., Bekaert, G., & Wei, M. (2007). Do macro variables, asset markets, or surveys forecast inflation better? *Journal of Monetary Economics*, *54*(4), 1163-1212. doi:10.1016/j.jmoneco.2006.04.006

Arias, J. E., Rubio-Ramirez, J. F., & Shin, M. (2022). Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models. *Journal of Econometrics*, *forthcoming*. doi:10.1016/j.jeconom.2022.04.013

Aruoba, S. B. (2020). Term structures of inflation expectations and real interest rates. *Journal of Business & Economic Statistics*, *38*(3), 542-553. doi:10.1080/07350015.2018.1529599

Banbura, M., Brenna, F., Paredes, J., & Ravazzolo, F. (2021). *Combining Bayesian VARs with survey density forecasts: Does it pay off?* (Working Paper Series No. 2543). European Central Bank.

Bassetti, F., Casarin, R., & Del Negro, M. (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 W. van der Klaauw, G. Topa, & R. Bachmann (Eds.), *Handbook of Economic Expectations.* Elsevier.

Bianchi, F., Ludvigson, S. C., & Ma, S. (2022). Belief distortions and macroeconomic fluctuations. *American Economic Review*, *112*(7), 2269-2315. doi:10.1257/aer.20201713

Boero, G., Smith, J., & Wallis, K. F. (2011). Scoring rules and survey density forecasts. *International Journal of Forecasting*, *27*(2), 379-393. doi:10.1016/j.ijforecast.2010.04.003

Cakmakli, C., & Demircan, H. (2022). Using survey information for improving the density nowcasting of US GDP. *Journal of Business & Economic Statistics*, *forthcoming*. doi:10.1080/07350015.2022.2058000

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business and Economic Statistics*, *29*(3), 327-341. doi:10.1198/jbes.2010.09248

Clark, T. E., Ganics, G., & Mertens, E. (2022, November). *Constructing fan charts from the ragged edge of SPF forecasts* (Working paper No. 22-36). Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202236

Clark, T. E., McCracken, M. W., & Mertens, E. (2020). Modeling time-varying uncertainty of multiple-horizon forecast errors. *The Review of Economics and Statistics*, *102*(1), 17-33. doi:10.1162/rest_a_00809

Clements, M. P. (2010). Explanations of the inconsistencies in survey respondents' forecasts. *European Economic Review*, *54*(4), 536–549. doi:10.1016/j.euroecorev.2009.10.003

Clements, M. P. (2014a). Forecast uncertainty–ex ante and ex post: U.S. inflation and output growth. *Journal of Business & Economic Statistics*, *32*(2), 206-216. doi:10.1080/07350015.2013.859618

Clements, M. P. (2014b). Probability distributions or point predictions? Survey forecasts of US output growth and inflation. *International Journal of Forecasting*, *30*(1), 99-117. doi:10.1016/j.ijforecast.2013.07.010

Clements, M. P. (2018). Are macroeconomic density forecasts informative? *International Journal of Forecasting*, *34*(2), 181-198. doi:10.1016/j.ijforecast.2017.10.004

Clements, M. P., & Galvão, A. B. (2017). Model and survey estimates of the term structure of US macroeconomic uncertainty. *International Journal of Forecasting*, *33*(3), 591-604. doi:10.1016/j.ijforecast.2017.01.004

Clements, M. P., Rich, R. W., & Tracy, J. (2022). Surveys of professionals. In W. van der Klaauw, G. Topa, & R. Bachmann (Eds.), *Handbook of Economic Expectations.* Elsevier.

Cogley, T., Morozov, S., & Sargent, T. J. (2005). Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control*, *29*(11), 1893-1925. doi:10.1016/j.jedc.2005.06.005

Coibion, O., & Gorodnichenko, Y. (2015). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, *105*(8), 2644-78. doi:10.1257/aer.20110306

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

Croushore, D., & Stark, T. (2019). Fifty years of the Survey of Professional Forecasters. *Economic Insights*, *4*(4), 1-11.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101. doi:10.1002/jae.1257

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253-63. doi:10.2307/1392185

Diebold, F. X., Shin, M., & Zhang, B. (2022). *On the aggregation of probability assessments: Regularized mixtures of predictive densities for Eurozone inflation and real interest rates* (NBER Working Papers No. 29635). National Bureau of Economic Research, Inc. doi:10.3386/w29635

Diebold, F. X., Tay, A. S., & Wallis, K. F. (1999). Evaluating density forecasts of inflation: The Survey of Professional Forecasters. In R. Engle & H. White (Eds.), *Cointegration, causality, and forecasting: A Festschrift in honour of Clive WJ Granger* (pp. 76–90). Oxford University Press.

Dieppe, A., Legrand, R., & van Roye, B. (2016). *The BEAR Toolbox* (Working Paper No. No. 1934). European Central Bank.

Engelberg, J., Manski, C. F., & Williams, J. (2009). Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, *27*(1), 30–41. doi:10.1198/jbes.2009.0003

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vols. 2, Part A, p. 2-56). Elsevier. doi:10.1016/B978-0-444-53683-9.00001-3

Galvão, A. B., Garratt, A., & Mitchell, J. (2021). Does judgment improve macroeconomic density forecasts? *International Journal of Forecasting*, *37*(3), 1247–1260. doi:10.1016/j.ijforecast.2021.02.007

Ganics, G., & Odendahl, F. (2021). Bayesian VAR forecasts, survey information, and structural change in the euro area. *International Journal of Forecasting*, *37*(2), 971–999. doi:10.1016/j.ijforecast.2020.11.001

Ganics, G., Rossi, B., & Sekhposyan, T. (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*.

Genre, V., Kenny, G., Meyler, A., & Timmermann, A. (2013). Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, *29*(1), 108-121. doi:10.1016/j.ijforecast.2012.06.004

Glas, A., & Hartmann, M. (2022). Uncertainty measures from partially rounded probabilistic forecast surveys. *Quantitative Economics*, *13*(3), 972-1022. doi:10.3982/QE1703

Grishchenko, O., Mouabbi, S., & Renne, J.-P. (2019). Measuring inflation anchoring and uncertainty: A U.S. and euro area comparison. *Journal of Money, Credit, and Banking*, *51*(5), 1053–1096. doi:10.1111/jmcb.12622

Jo, S., & Sekkel, R. (2019). Macroeconomic uncertainty through the lens of professional forecasters. *Journal of Business & Economic Statistics*, *37*(3), 436-446. doi:10.1080/07350015.2017.1356729

Krüger, F. (2017). Survey-based forecast distributions for Euro area growth and inflation: ensembles versus histograms. *Empirical Economics*, *53*(1), 235-246. doi:10.1007/s00181-017-1228-3

Krüger, F., Clark, T. E., & Ravazzolo, F. (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business & Economic Statistics*, *35*(3), 470-485. doi:10.1080/07350015.2015.1087856

Krüger, F., & Pavlova, L. (2022). *Quantifying subjective uncertainty in survey expectations.* (*mimeo*, Karlsruhe Institute of Technology)

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427-443. doi:10.1002/jae.695

Mertens, E. (2016). Measuring the level and uncertainty of trend inflation. *The Review of Economics and Statistics*, *98*(5), 950-967. doi:10.1162/REST_a_00549

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

Patton, A. J., & Timmermann, A. (2011). Predictability of output growth and inflation: A multi-horizon survey approach. *Journal of Business & Economic Statistics*, *29*(3), 397-410. doi:10.1198/jbes.2010.08347

Patton, A. J., & Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, *30*(1), 1-17. doi:10.1080/07350015.2012.634337

Robertson, J. C., Tallman, E. W., & Whiteman, C. H. (2005). Forecasting using relative entropy. *Journal of Money, Credit and Banking*, *37*(3), 383-401. doi:10.1353/mcb.2005.0034

Tallman, E. W., & Zaman, S. (2020). Combining survey long-run forecasts and nowcasts with BVAR forecasts using relative entropy. *International Journal of Forecasting*, *36*(2), 373-398. doi:10.1016/j.ijforecast.2019.04.024

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
GDP price index	PGDP	1968Q4	1981Q3	NA	NA
		Panel B: Histograms			
Real GDP growth	PRGDP	NA	1981Q3	2009Q2	2009Q2
GDP price index inflation	PRPGDP	NA	1981Q3	NA	NA

Note: The table reports the first quarters for which 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) that are, however, disregarded in our analysis. In addition, due to potential data issues (discussed in the text), we use only histogram data as of 1992.

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

	SV		Relative to SV (in denominator)					
			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	1.98	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	1.01	1.00	0.99
6	5.06	2.28	1.00	0.99	1.00	1.02**	1.01	1.01
7	5.09	2.33	1.00	0.99	1.00	1.01	1.00	1.00
8	5.12	2.29	1.00	0.99	1.00	1.01	1.00	1.00
9	5.10	2.24	1.00	0.99	1.00	1.02	1.00	1.01
10	5.12	2.20	1.00	1.00	1.00	1.00	1.00	1.00
11	5.17	2.22	1.00	0.99	1.00	1.01	0.99	0.99
12	5.19	2.21	1.00	0.99	1.00	1.01	0.99	1.00
13	5.20	2.32	1.00	1.00	1.00	1.01	1.00	1.01
14	5.23	4.61	1.00	1.00	1.00	1.00	1.00	1.00
15	5.26	5.33	1.00	1.00	1.01	1.01	1.01	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*	1.02	1.04*	1.02
2	1.69	1.15	0.99	1.00	1.00	1.02	1.01	1.01
3	1.73	1.19	0.99	0.99	1.01	1.02	1.01	1.01
4	1.77	1.21	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	1.02
6	1.82	1.23	0.99	0.99	1.03	1.06***	1.02	1.05*
7	1.86	1.26	1.00	0.99	1.03	1.05**	1.02	1.04
8	1.86	1.24	0.99	0.99	1.03	1.05**	1.02	1.04
9	1.83	1.21	1.00	0.99	1.04**	1.07***	1.03**	1.05*
10	1.83	1.19	1.00	0.99	1.03*	1.06***	1.03*	1.05*
11	1.85	1.20	0.99	0.98	1.03**	1.07***	1.02	1.04
12	1.86	1.20	0.99	0.99	1.03**	1.07***	1.03*	1.05**
13	1.87	1.25	1.00	1.00	1.04**	1.06**	1.03**	1.06**
14	1.89	1.64	1.00	1.00	1.04**	1.05**	1.04**	1.05**
15	1.89	1.88	1.00	1.00	1.05***	1.05***	1.04***	1.05***

Note: Forecasts for quarterly GDP growth h steps ahead over subsamples extending from 1992Q1 until 2022Q2 and 2016Q4, respectively (using data for realized values as far as available in 2022Q2). Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Table 3: Accuracy of forecasts of GDP price inflation with and without entropic tilting

			Relative to SV (in denominator)					
	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	0.93	0.76	1.01*	1.01	1.00	1.00	1.01*	1.02
1	1.10	0.84	1.01*	1.03**	1.00	1.00	1.01	1.03*
2	1.15	0.88	1.02**	1.04**	1.00	1.00	1.03**	1.04*
3	1.19	0.91	1.02	1.04*	1.00	1.00	1.02*	1.05*
4	1.26	0.99	1.03**	1.05**	1.00	1.00	1.03**	1.05**
5	1.28	1.07	1.02*	1.03*	1.00	1.01	1.02	1.04
6	1.26	0.99	1.02*	1.04**	1.03***	1.04***	1.04***	1.07***
7	1.24	0.97	1.02*	1.03*	1.03	1.06*	1.04**	1.08***
PANEL B: CRPS								
0	0.49	0.43	1.01	1.01	1.05***	1.04*	1.05***	1.05**
1	0.58	0.48	1.02*	1.03*	1.04***	1.05**	1.05***	1.07***
2	0.61	0.51	1.03**	1.03**	1.04**	1.05***	1.06***	1.07***
3	0.63	0.53	1.03**	1.04**	1.05***	1.06***	1.06***	1.08***
4	0.67	0.58	1.03**	1.04**	1.04**	1.04**	1.05***	1.07***
5	0.71	0.62	1.02*	1.03*	1.06***	1.08***	1.06***	1.09***
6	0.69	0.59	1.02**	1.03**	1.12***	1.17***	1.12***	1.18***
7	0.69	0.58	1.02**	1.03**	1.15***	1.22***	1.15***	1.23***

Note: Forecasts for quarterly inflation h steps ahead over subsamples extending from 1992Q1 until 2022Q2 and 2016Q4, respectively (using data for realized values as far as available in 2022Q2). Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

Table 4: Relative forecast accuracy when tilting to fitted moments instead of histograms

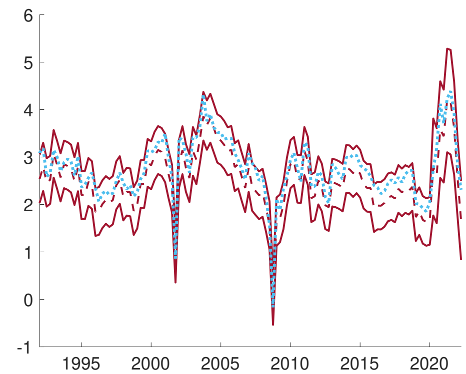
	RMSE				CRPS			
	SV		CONST		SV		CONST	
h	92–22	92–16	92–22	92–16	92–22	92–16	92–22	92–16
Panel A: GDP growth								
0	1.01*	1.00	1.00	1.00	1.01***	1.01***	1.00	1.00
1	1.00	1.00	1.00	1.00	1.01	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
10	1.00	1.00	1.00	1.00	1.00	0.99**	1.00	1.00
11	1.00	1.00	1.00	1.00	1.00	0.99***	1.00	1.00
12	1.00	1.00	1.00	1.00	1.00	0.99***	1.00	1.00
13	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
14	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
15	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Inflation in the GDP price index								
0	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01*	1.01*
1	1.01	1.01	1.01	1.01*	1.01	1.01	1.00	1.00
2	1.01**	1.02**	1.01**	1.02**	1.01*	1.02*	1.01	1.01
3	1.01***	1.03***	1.01***	1.02***	1.01**	1.02*	1.01	1.01
4	1.02***	1.03***	1.01**	1.02***	1.01**	1.02*	1.01	1.01
5	1.01**	1.02***	1.01**	1.01**	1.01*	1.01	1.00	1.00
6	1.01**	1.01**	1.00	1.01*	1.01*	1.01	1.00	1.00
7	1.01*	1.01*	1.00	1.00	1.00	1.00	1.00	1.00

Note: Relative RMSE and CRPS of forecasts obtained from tilting to moments imputed from the SPF probability bins vs. tilting to the bins directly (with the latter in the denominator). The imputed moments are mean, variance, and skewness of a generalized beta distribution fitted to the SPF probability bins. Forecasts are for quarterly outcomes and evaluation windows extend from 1992Q1 until 2022Q2 and 2016Q4, respectively (using data for realized values as far as available in 2022Q2). Significance assessed by Diebold-Mariano tests using Newey-West standard errors with $h+1$ lags. ***, ** and * denote significance at the 1%, 5%, and 10% level, respectively.

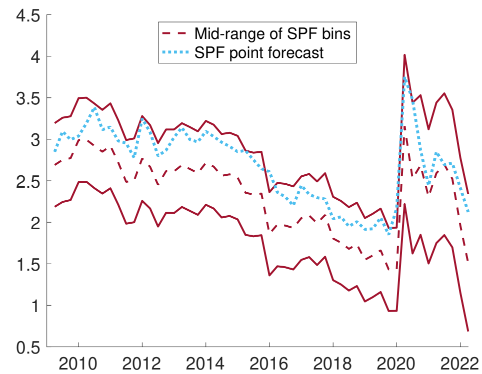
Figure 1: SPF annual point forecasts compared to means implied by SPF probability bins

Next-year forecasts | **2- and 3-years-ahead forecasts**

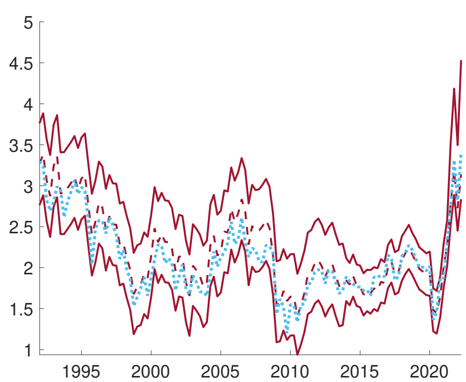
(a) GDP growth



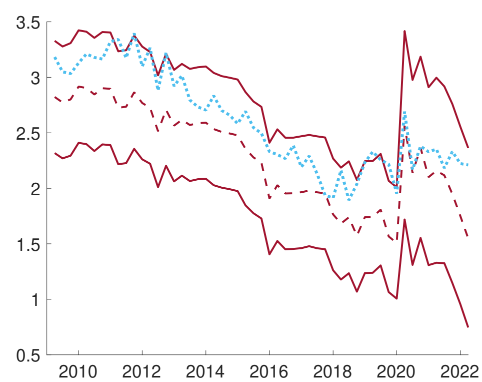
(b) GDP growth, 2-years-ahead



(c) GDP price index inflation

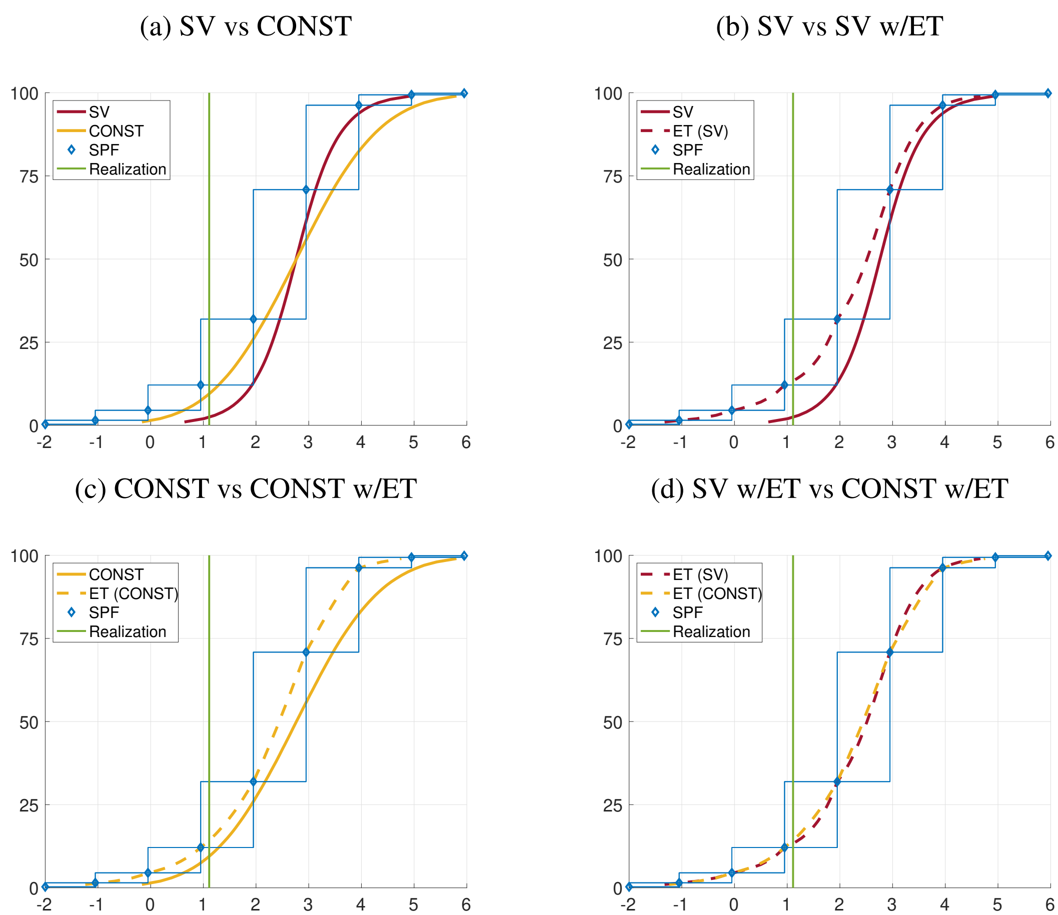


(d) GDP growth, 3-years-ahead



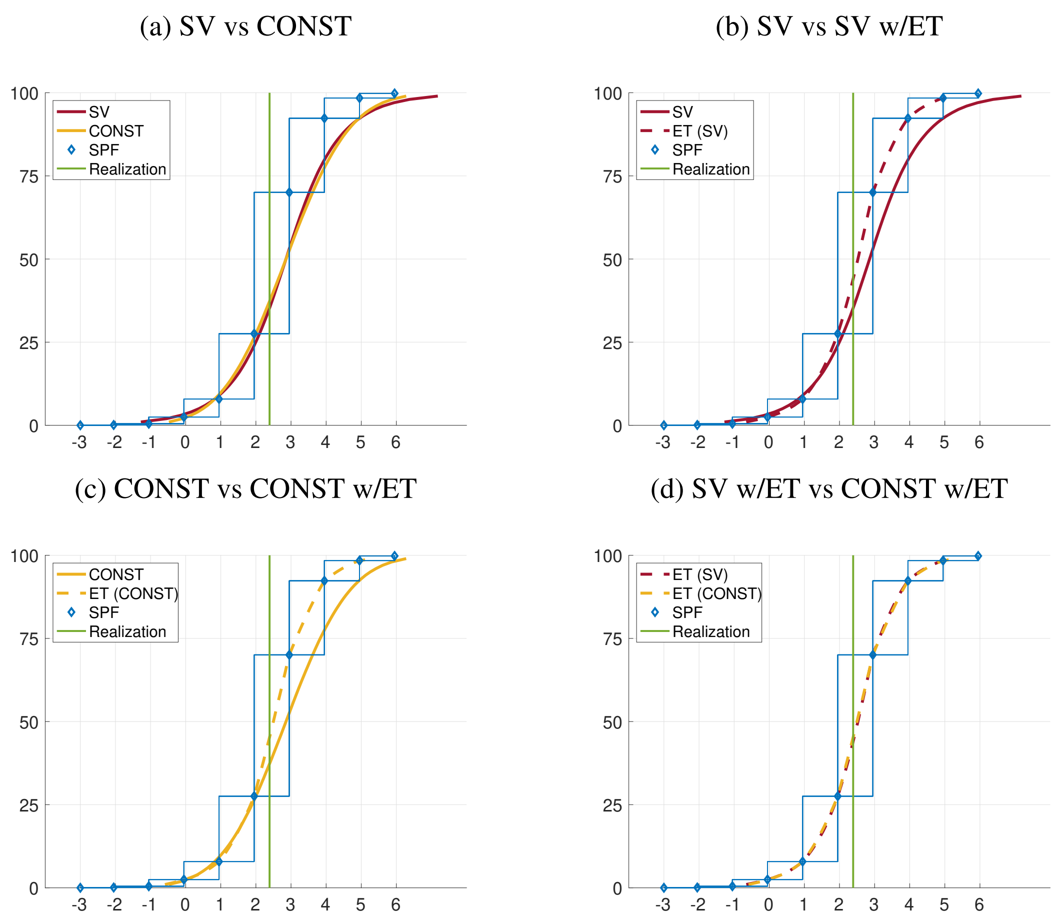
Notes: The panels compare SPF point forecasts (dotted blue lines) against lower and upper bounds of the range of mean forecasts consistent with the histograms (solid red lines) along with the corresponding mid-range values (dashed red lines).

Figure 2: Cumulative distribution functions for GDP growth with and without tilting, in 2007Q3



Notes: Cumulative distributions obtained from the SV and CONST models (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 lines). Probabilities on the y-axis in percentage points, and values for annualized GDP growth in percent on the x-axis.

Figure 3: Cumulative distribution 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

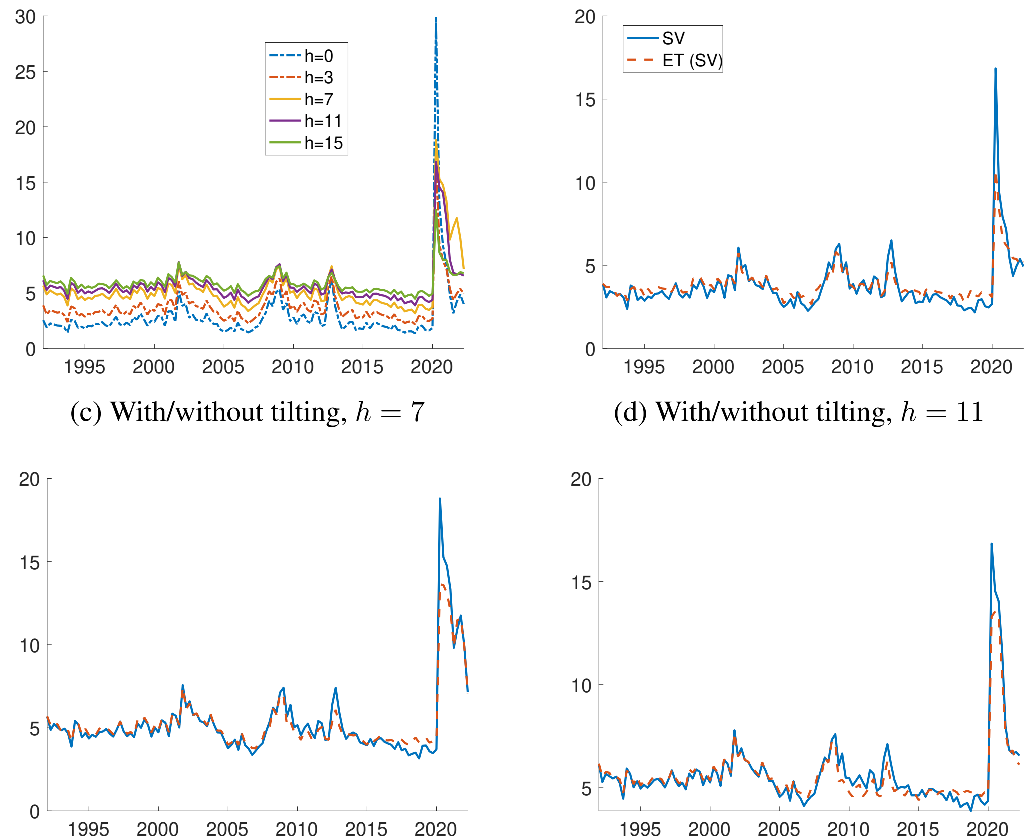
(d) SV w/ET vs CONST w/ET

Notes: Cumulative distributions obtained from the SV and CONST models (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 lines). Probabilities on the y-axis in percentage points, and values for annualized GDP growth in percent on the x-axis.

Figure 4: Term structures of GDP growth uncertainty

(a) Model-based estimates across horizons

(b) With/without tilting, $h = 3$

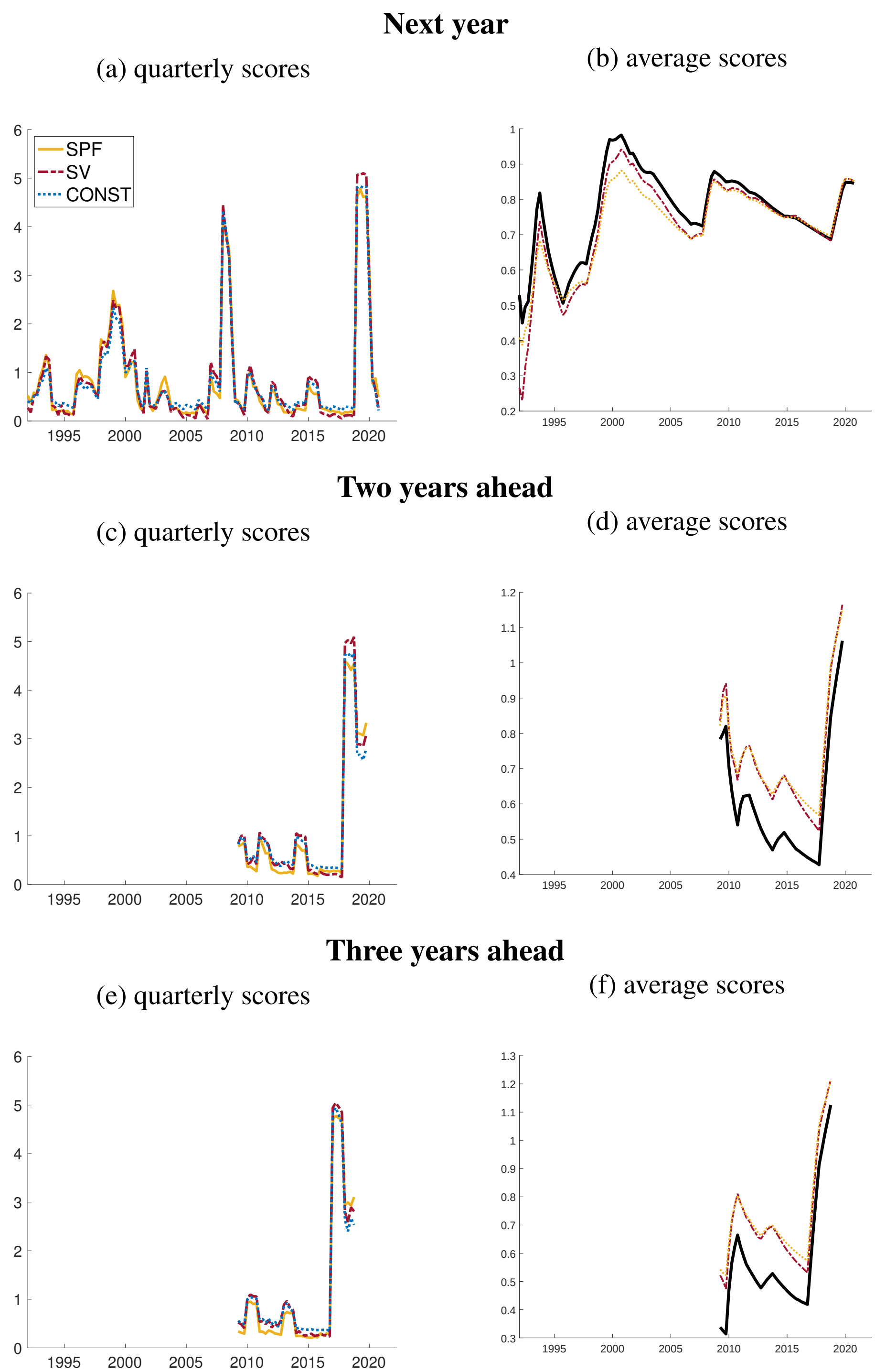


(c) With/without tilting, $h = 7$

(d) With/without tilting, $h = 11$

Notes: Term structure of GDP growth uncertainty, measured by the width of 68% bands of predictive real-time densities, SV model with and without entropic tilting to the SPF histograms. In panels (b) through (d), the solid blue lines provide model-based estimates, and the dashed red lines correspond to entropically tilted estimates.

Figure 5: DRPS of models and SPF, for annual forecasts of GDP growth



Notes: DRPS 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 or 2009Q2 in the case of next-year or two- and three-year-ahead horizons, respectively.