Entropic Tilting of Forecasts to SPF Histograms: Analytics & Applications*

Todd E. Clark,[1] Elmar Mertens[2]

[1]*Johns Hopkins University and Federal Reserve Bank of Cleveland,* [2]*European Central Bank.*

March 22, 2026

Abstract

This paper develops a new, direct approach to entropic tilting of model-based predictive distributions to match histogram forecasts provided in the U.S. Survey of Professional Forecasters (SPF). We focus on tilting to histogram probabilities directly, rather than to moments of fitted distributions. We reformulate the single-histogram tilting problem and derive a novel analytic characterization for the multiple-histogram case, with iterative solutions via Iterative Proportional Fitting. Application to quarterly real-time forecasts of major macroeconomic aggregates from a Bayesian vector autoregression with time-varying volatility shows that tilting to SPF histograms significantly improves on the model's baseline forecasts, particularly during periods around the Great Recession and the COVID-19 pandemic. Crucially, entropic tilting not only improves accuracy for those model variables for which histogram targets are available but also for other model variables for which SPF targets are not available.

*We gratefully acknowledge helpful suggestions and comments received from Dean Croushore, Domenico Giannone, Michele Lenza, Simon van Norden, Bernd Schwaab, and conference and seminar participants at the Real Time Econometrics conference 2025 (Eltville), ECB, and University of Pittsburgh. The views expressed herein are solely those of the authors and do not necessarily reflect the views of the European Central Bank, the Eurosystem, the Federal Reserve Bank of Cleveland, or the Federal Reserve System.

1 Introduction

An immense literature has explored the predictive content of forecasts from the U.S. Survey of Professional Forecasters (SPF) or its counterpart in the euro area. Within this body of work, many studies have focused on survey point forecasts, resulting in these forecasts being widely recognized for their real-time accuracy. While less studied, the SPF (both in the U.S. and euro area) also includes density forecasts, expressed as histograms over forecast bins. The merits of these density forecasts have been more widely debated. Some research has found value in these projections. For example, Clements (2004, 2014) finds that such histograms can offer informative density forecasts, often competitive with or superior to parametric model-based alternatives during periods of heightened uncertainty. On the other hand, studies such as Clements (2008) and Engelberg, Manski, and Williams (2009) have noted several limitations of histogram forecasts, including issues of coherence, underdispersion, and resolution due to fixed binning and finite responses.[1] However, it should be noted that some of these issues are more prevalent in individual forecasters' histograms than in the aggregate histograms that average across forecasters.

One of the ways that survey forecasts have been explored or used is to try to improve forecasts from time series models, such as vector autoregressions, by combining the alternative projections. For example, Faust and Wright (2013) use current-quarter point forecasts from surveys as jumping off points to enhance model-based forecasts. More formally, point forecasts from surveys can be brought to inform model forecasts through (Gaussian) conditional forecasting methods, with model forecasts conditioning on survey forecasts. A closely related, less parametric approach is to use the method of entropic tilting introduced into economic forecasting by Robertson, Tallman, and Whiteman (2005). In the tilting approach, the model's predictive distribution is modified so that the relevant moments of the modified distribution match the chosen moments from the survey forecasts. Our

[1] Relatedly, see also Knüppel and Pavlova (2023). In addition, Lenza, Moutachaker, and Paredes (2025) develop a quantile regression forecast model that improves — at short horizons — on the density forecast accuracy of the ECB's SPF for the euro area.

paper rather emphasizes the application of entropic tilting to directly match probabilistic forecasts, as opposed to matching the moments of an underlying distribution.

Most studies that entropically tilt model-based forecasts toward SPF information focus on a survey's point forecasts, and interpret these as targets for matching the predictive mean. As examples, Krueger, Clark, and Ravazzolo (2017) entropically tilt forecasts from a Bayesian vector autoregression (BVAR) to (point) nowcasts from the SPF, and Tallman and Zaman (2020) tilt BVAR forecasts to (point) long-run survey forecasts.[2] The existing work that uses density forecasts from the SPF to inform model-based forecasts does so by extracting a limited set of moments from the SPF histograms. Galvão, Garratt, and Mitchell (2021) and Banbura, et al. (2024) proceed by fitting continuous distributions (parametrically and non-parametrically, respectively) to summarize histogram information in the form of first or second moments from the fitted distributions.[3] Ganics and Odendahl (2021) impute first and second moments under the assumption that all mass within a given histogram bin is concentrated at the center of the bin. All of this work on combining survey and model forecasts uses aggregate (average or median) forecasts from the SPF.

This paper develops and examines a different, direct approach to combining information in density forecasts of the SPF with model-based forecasts. Our approach takes the SPF histograms as given and avoids the need to fit distributions to the histograms and rely on moments from the fitted distribution. Accordingly, our approach does not require taking a stand on how the SPF forecaster distributes predictive mass within each histogram bin or taking a stand on a parametric or non-parametric distribution to characterize the histogram and then in turn estimating the distribution. Instead, our approach directly tilts the predictive distribution of the BVAR to the SPF histogram bins, using information directly in the SPF and nothing more to set the tilting targets. Throughout, we focus

[2]Additional applications of entropic tilting to BVAR forecasts include Cogley, Morozov, and Sargent (2005), and Bobeica and Hartwig (2023).

[3]Specifically, Galvão, Garratt, and Mitchell (2021) extract moment parameters from fitted generalized beta distributions (Engelberg, Manski, and Williams, 2009), whereas Banbura, et al. (2024) employ random sampling from multinomial distributions and kernel-smoothed estimates from the resulting densities.

our application to histograms of the average SPF forecaster, also known as "consensus" histograms.[4]

Our paper has two primary contributions. The first is methodological: We develop a new analytical characterization for the tilting problem when targeting histogram bins. As noted already by Tallman and West (2022), the tilting solution is particularly simple when only a single histogram is targeted, and reweighs model draws within each bin to match SPF-provided bin probabilities. Compared to their work, we reformulate the solution in the single-histogram case, and extend the solution to the case of multiple histograms. In addition, we discuss implementation details.

Our second contribution is empirical: Focusing on the U.S., we examine the information content of SPF histogram forecasts for the accuracy of forecasts of major macroeconomic aggregates produced by a Bayesian vector autoregression. Using our analytical solution for the tilting problem, we entropically tilt the BVAR's quarterly forecast distributions to SPF histograms, using the available multi-year forecast horizons from the SPF for three variables, including GDP growth, the unemployment rate, and core PCE inflation.[5] The BVAR includes stochastic volatility with fat-tailed innovations to the VAR and volatility outliers, a specification recently developed by Carriero, et al. (2024) and found to forecast well in a long historical sample, including the high volatility induced by the COVID-19 pandemic. We provide evidence that, relative to the baseline model forecasts, SPF histogram forecasts add the most information content around the Great Recession and its aftermath, as well as during the COVID-19 pandemic, with most of the information coming from the SPF's GDP growth and unemployment rate forecasts. As our baseline already allows for some flexibility in adapting to changing economic conditions, in particular during times of turbulence, the gains from tilting densities from less flexible models are, or

[4]For a discussion of potential difficulties in handling individual SPF histograms see, for example, Bassetti, Casarin, and Del Negro (2023) and Clark and Mertens (2024).

[5]Our development and use of an analytical solution to the problem of tilting to histograms and application to BVAR forecasts distinguishes this paper from the earlier work of Clark, Ganics, and Mertens (2022), which instead relied on the standard numerical solution of the tilting problem and tilted SPF-consistent point forecasts obtained from a model specified in state-space form to the SPF histograms.

course, larger.[6]

Our examination of the forecast accuracy between the tilted and the untilted distribution of the BVAR shows that, on average over the period from 1996 through 2025, entropically tilting BVAR forecasts to the annual SPF histograms consistently improves the accuracy of forecasts for quarterly variables. The size of these forecast gains varies by horizons and across variables. Crucially, entropic tilting not only improves accuracy for those model variables for which histogram targets are available but also the other model variables (e.g., employment) for which targets are not available. Drilling into accuracy over time shows that the benefits of entropic tilting BVAR-based forecasts to SPF forecast histograms are driven by the period since 2007 (i.e., since the Great Recession), perhaps related to the fact that the SPF began publication of unemployment rate forecasts and longer-horizon forecasts for growth only in 2009, and Core PCE inflation forecasts have been published only since 2007.[7] While helpful to forecast accuracy on average for the 2007-2025 sample, the benefits of entropically tilting the BVAR forecasts to the SPF histograms are sharpest in the years around the Great Recession and around COVID's outbreak in 2020.

The paper proceeds as follows. Section 2 details the connection of our approach to other work on the method of entropic tilting. Section 3 develops our analytical solution to tilting to histogram forecasts. Section 4 describes the data used and details the forecasting BVAR. Section 5 presents empirical results of tilting BVAR forecasts to SPF density forecasts. Section 6 concludes. Additional details and results are provided in a supplementary online appendix (pending).

[6]Corresponding results for tilting a standard Gaussian BVAR are available upon request. See also the results of Bobeica and Hartwig (2023) related to robustness of BVARs tilted to the ECB's SPF during the COVID-19 pandemic.

[7]Relatedly, the SPF has collected probabilistic forecasts for core PCE inflation only since 2007.

2 Related literature

Our use of entropic tilting to align model-implied forecast densities with survey histograms particularly across multiple forecast horizons can be viewed as a modern and applied extension of the classical *Iterative Proportional Fitting* (IPF) algorithm. In our methodology, IPF serves to iteratively reweight a joint distribution so that its marginals match externally imposed constraints (e.g., SPF histogram bins), while minimizing the Kullback-Leibler (KL) divergence from a prior distribution. This structure of iterative KL projection is rooted in foundational work on IPF (Deming and Stephan, 1940), later formalized as an I-projection by Csiszár (1975) and given geometric interpretation by Fienberg (1970). Our approach also connects naturally to the literature on empirical likelihood and exponential tilting. Schennach (2007) demonstrates how exponential tilting arises in moment-based estimation problems as a dual representation of KL minimization.

A related strand develops the theoretical and structural underpinnings of exponential-tilting adjustments. Giacomini and Ragusa (2014) apply the same Kullback-Leibler-projection principle to impose nonlinear theoretical moment conditions, such as Euler-equation restrictions, on model-based forecast densities. Moreover, Giacomini and Ragusa (2014) establish that entropic tilting improves forecast accuracy (as measured by logarithmic scores) when the tilting targets are true in population. Their "theory-coherent forecasting" approach and our framework share identical entropic geometry but differ in purpose: theirs enforces structural coherence with economic theory, ours empirical coherence with survey-implied probability mass functions. Antolín-Díaz, Petrella, and Rubio-Ramírez (2021) further show that conditional-forecasting and structural-scenario analyses in SVARs can likewise be derived as KL-minimizing projections, providing a unifying information-theoretic view of constrained forecasting across theoretical, structural, and empirical contexts.

Our methodology builds directly on and extends the literature on entropic tilting and constrained Bayesian forecasting, including key contributions by Mike West and coauthors.

In particular, Tallman and West (2022) and West (2024) develop a foundational treatment of entropic tilting as a Bayesian decision-analytic method for modifying forecast distributions under expectation constraints, including cases with quantile or indicator function targets. Their analytic treatment of tilting under histogram- or quantile-based constraints shares the same structure as our own KL projection approach, though our implementation emphasizes iterative reweighting across multiple marginals, leveraging computational tools from iterative proportional fitting. The related work of West (2024) further generalizes constrained forecasting within a decision-theoretic framework, advocating entropic tilting as a principled alternative to naive probabilistic conditioning. This distinction is particularly relevant in our context, where survey histograms represent external, judgment-based constraints not generated within the probabilistic system of the model itself. Our approach complements the scenario synthesis framework of Adrian, et al. (2025), which employs entropic tilting to reconcile a baseline forecast with externally specified scenarios. While their emphasis lies in evaluating the concordance of narrative scenarios with a statistical reference using expected misclassification rates, our contribution focuses on reconciling high-frequency density forecasts with survey-based marginal distributions for multiple variables and multiple forecast horizons.

In a related but distinct contribution, Montes-Galdón, Paredes, and Wolf (2026) propose "parametric tilting" as an alternative to standard entropic tilting that addresses robustness concerns arising when moment-based tilting targets lie far from the support of the baseline density. Rather than reweighting draws directly, their approach minimizes the Kullback-Leibler divergence between the model-based density and a skew-t distribution subject to moment constraints on, e.g., the mean, mode, variance, or skewness of the target distribution. By construction, the resulting tilted density inherits the parametric regularity of the skew-t family, guaranteeing unimodality and numerical stability even when the target moments deviate substantially from those of the baseline—a setting in which standard entropic tilting can produce degenerate or multimodal distributions, as illustrated by

Banbura, et al. (2024) during the COVID-19 pandemic. Our approach is complementary: whereas parametric tilting operates on a small number of moment constraints and imposes a parametric shape on the tilted density, our method works directly with histogram bin probabilities and preserves the nonparametric flexibility of the baseline density within each bin. This distinction is empirically relevant, since our results in Section 5.5 indicate that the full histogram contains predictive information beyond what can be captured by a few summary moments.

More broadly, our paper also contributes to the literature that evaluates the predictive content of probabilistic survey forecasts. A large part of this literature has examined heterogeneity, consistency, and performance characteristics across individual respondents.[8] By contrast, and similar to the literature surveyed in the previous section, we investigate the predictive value of aggregated probability forecasts, where issues of coherence and consistency are fairly negligible, as we document below.

Our focus on directly tilting model-based forecasts to match histogram probabilities complements this literature by providing a practical method to leverage the information content of survey histograms without imposing additional parametric assumptions. As such, our out-of-sample analysis differs from that of Giacomini and Ragusa (2014), who evaluate entropic tilting as a means of imposing theory-coherent moment restrictions—such as Euler equations—on model-based predictive densities. Their exercise tests whether such structural coherence improves forecast performance in a frequentist rolling-window setting. By contrast, we evaluate empirical coherence: our tilting aligns Bayesian predictive densities with survey-implied probability distributions from the SPF and assesses the resulting forecasts using standard density-forecast scoring rules. Thus, whereas Giacomini and Ragusa's analysis validates theoretical consistency, our framework tests calibration gains from survey consistency within a Bayesian, analytically tractable implementation.

[8] See, for example, Engelberg, Manski, and Williams (2009), Clements (2010), Rich, Song, and Tracy (2012), Kenny, Kostka, and Masera (2014), Kenny, Kostka, and Masera (2015), Meyler (2020), Bassetti, Casarin, and Del Negro (2022, 2023), Coroneo, Iacone, and Profumo (2024), Allayioti, et al. (2024), Clements, Rich, and Tracy (2025), and Clements (2025).

By evaluating the forecast performance of tilted distributions, we contribute empirical evidence on the value of survey-based density information in enhancing model-based forecasts of key macroeconomic variables. Moreover, we document the added value of survey histograms particularly during periods of economic turbulence, such as the Great Recession and the COVID-19 pandemic, relative to state-of-the-art time-series models highlighting the practical relevance of our approach for real-world forecasting challenges.

3 Analytical tilting to histograms

Here, we present our new analytical solution to the problem of tilting to histograms. We begin with describing the basic setup of the tilting problem. We then take up the solution to the problem. Given its particular simplicity, we first discuss the case of tilting to a single histogram. We then describe the generalization to two and more histograms. In addition, we discuss our computational implementation of the tilting solution in the occasional cases in which the original predictive density from the model does not have mass in some of the target bins from SPF histogram forecasts.

3.1 Setup of the tilting problem

We seek to modify the predictive density obtained from a model so that it matches the histogram probabilities from the SPF, while otherwise minimizing the relative entropy (or Kullback-Leibler divergence) between the modified density and the original BVAR-based predictive density. This is the problem of entropic tilting to match histogram forecasts. For concreteness, and foreshadowing our empirical application, we will refer to the model as a BVAR. Since entropic tilting operates only on draws generated from the model's predictive density (without regard for its underlying structure), it can generally be applied to models other than a BVAR, such as (non-linear) DSGE models, QVARs etc. as well.

Formally, let y_t denote the vector of variables of interest, such as GDP growth, inflation

and the unemployment rate. At the forecast origin of period t, the estimated model provides a baseline predictive density for future outcomes for horizons $h = 1, 2, \ldots, H$. We denote the vector of future outcomes at these horizons as $\boldsymbol{X}_t = (y_{t+1}, y_{t+2}, \ldots, y_{t+H})$, and the BVAR's predictive density for $\boldsymbol{X}_t$ as $p_t(\boldsymbol{X}_t)$.

The SPF provides histogram predictions for calendar-year outcomes with data definitions that vary from variable to variable. In general, the outcomes considered by the SPF are non-linear transformations of $\boldsymbol{X_t}$. For SPF histogram n, we denote the applicable transformation at round t by $f_t^n(\boldsymbol{X_t})$.[9] Different histograms may convey predictions for different variables, such as GDP growth and the unemployment rate, and/or predictions for the same variable but at different forecast horizons, such as one-year-ahead and two-year-ahead forecasts of GDP growth. In general, reflecting the position of the current quarter t in the calendar year, the transformations $f_t^n(\cdot)$ are time-dependent, may involve multiple (though typically not all) elements of $\boldsymbol{X}_t$, and can be non-linear. For example, for the unemployment rate, $f_t^n(\boldsymbol{X}_t)$ is the linear average of the quarterly unemployment rates for the targeted year contained in $\boldsymbol{X}_t$. However, for GDP growth and PCE (core) inflation the corresponding definitions of $f_t^n(\boldsymbol{X}_t)$ are non-linear transformations of $\boldsymbol{X_t}$. Specifically, for GDP growth, $f_t^n(\boldsymbol{X}_t)$ is the simple percentage change in the annual average level of real GDP, which involves the ratio of average annual levels of real GDP between the targeted and the previous year. For PCE core inflation, $f_t^n(\boldsymbol{X}_t)$ corresponds to the Q4 over Q4 change in the level of the PCE core price index. A further source of non-linearity in the definitions of $f_t^n(\boldsymbol{X}_t)$ is typically also that $\boldsymbol{X_t}$ consists of differences in log levels (rather than simple growth rates).

Let y_t^{GDP}, y_t^{UNRATE}, and y_t^{INF} denote the elements of y_t that correspond to the quarterly values of log-differenced real GDP, the unemployment rate, and log-differenced core PCE inflation, respectively. Considering predictions for the calendar year comprising the quarters $t+h-3$ through $t+h$, we obtain the following definitions for the transformations

[9] We index the SPF histograms by $n = 1, 2, \ldots, N$, where N is the number of histograms to which we seek to tilt the BVAR's predictive density.

$f_t^n(\cdot)$ for the three variables of interest:

$$\text{Unemployment:} \quad f_t^n(\boldsymbol{X}_t) = \frac{1}{4} \cdot \sum_{i=h-3}^{h} y_{t+i}^{\text{UNRATE}} \tag{1}$$

$$\text{GDP growth:} \quad f_t^n(\boldsymbol{X}_t) = 100 \times \frac{Y_{t+h-3}^{\text{GDP}} + Y_{t+h-2}^{\text{GDP}} + Y_{t+h-1}^{\text{GDP}} + Y_{t+h}^{\text{GDP}}}{1 + Y_{t+h-6}^{\text{GDP}} + Y_{t+h-5}^{\text{GDP}} + Y_{t+h-4}^{\text{GDP}}} - 100 \tag{2}$$

$$\text{Inflation:} \quad f_t^n(\boldsymbol{X}_t) = 100 \times \exp\left(\sum_{i=h-3}^{h} y_{t+i}^{\text{INF}}\right) - 100 \tag{3}$$

where $Y_{t+h}^{\text{GDP}} \equiv \exp\left(\sum_{i=h-6}^{h} y_{t+i}^{\text{GDP}}\right)$ denotes the quarterly level of GDP in $t+h$ relative to its level in $t+h-7$. For $i < 1$ the values of y_{t+i}^{UNRATE}, y_{t+i}^{GDP}, and y_{t+i}^{INF} reflect fixed and known data points, whereas for $i \geq 1$ their values are generated by the model .

As detailed with examples in Section 4, the SPF collects and publishes forecasts of probability distributions in the form of histograms with pre-specified bins of outcome ranges. More formally, let each histogram have J bins, and we can say that the SPF collects and publishes

$$\widehat{\text{Prob}}_t\left(b_{j-1}^n < f_t^n\left(\boldsymbol{X}_t\right) \leq b_j^n\right) \forall\ j, n, \tag{4}$$

where $b_{t,j}^n$ denotes the upper limit for the j bin, the lower limit of the first bin is $b_{t,0}^n$, and open-ended bins are represented as $b_{t,0}^n = -\infty$ and $b_{t,J}^n = \infty$.[10] To simplify notation, let $A_{t,j}^n$ denote the set of outcomes inside the j-th bin of histogram n:

$$A_{t,j}^n \equiv \{\boldsymbol{X}_t\ :\ b_{t,j-1}^n < f_t^n\left(\boldsymbol{X}_t\right) \leq b_{t,j}^n\}, \tag{5}$$

and we can refer to the SPF probabilities as $\widehat{\text{Prob}}_t\left(\boldsymbol{X}_t \in A_{t,j}^n\right)$.

Henceforth, for simplicity we drop the t subscripts on the set of outcomes, bin limits,

[10]For simplicity, our notation treats all histograms as having an identical number of J bins. Our implementation recognizes that the number of bins can vary across histograms, and the number of bins of histogram n is more generally denoted $J(n)$.

$\boldsymbol{X}$, and the transformation function $f^n(\cdot)$. Each histogram provides probabilities for a set of J mutually exclusive and exhaustive events $\{A_j^n\}_{j=1}^J$. For the baseline density, $p(A_j^n) \equiv p(\boldsymbol{X} \in A_j^n)$ describes the probabilities for $\boldsymbol{X}$ to fall in the jth bin of the histogram. Since the events are mutually exclusive and exhaustive we have $\sum_{j=1}^J p(A_j^n) = 1$. The tilted density $q(\boldsymbol{X})$ is required to match the histogram probabilities, so that $q(A_j^n) \equiv q(\boldsymbol{X} \in A_j^n)$ are the target probabilities from the SPF histogram, with $\sum_{j=1}^J q(A_j^n) = 1$. We seek to construct a tilted density, $q(\boldsymbol{X})$, that minimizes the relative entropy (or Kullback-Leibler divergence measure)

$$D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) = E_x^q \, \log\left(\frac{q(\boldsymbol{X})}{p(\boldsymbol{X})}\right) = \int_x q(\boldsymbol{X}) \, \log\left(\frac{q(\boldsymbol{X})}{p(\boldsymbol{X})}\right) dx \tag{6}$$

subject to matching histogram probabilities. As discussed, for example, by Cover and Thomas (2006), relative entropy is non-negative, $D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) \geq 0$, and equal to zero if and only if $p(\boldsymbol{X}) = q(\boldsymbol{X})$. Formally, the tilting problem can be expressed as

$$\min_q D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) \quad \text{s.t.} \quad q\big(\boldsymbol{X} \in A_i^n\big) = \widehat{\text{Prob}}\big(\boldsymbol{X} \in A_i^n\big) \quad \forall\, i, n. \tag{7}$$

The tilting problem in (7) is a special case of the original entropic tilting problem described by Robertson, Tallman, and Whiteman (2005), who considered a more general set of constraints, expressed as moment conditions. Since the event probabilities $q\big(\boldsymbol{X} \in A_i^n\big) = E\left(\mathbb{1}\left(\boldsymbol{X} \in A_i^n\right)\right)$ are particular moments of the distribution $q(\boldsymbol{X})$, the problem in (7) is a special case of their setup. As discussed by Robertson, Tallman, and Whiteman (2005), the entropic tilting problem is a convex optimization problem, and the solution is unique. In the next two subsections we present an analytical solution for the single-histogram case, and a novel analytical characterization for the multiple-histogram case, with iterative solution via Iterative Proportional Fitting.

For the single-histogram case, our solution is a mere reformulation of the solution described by Tallman and West (2022) and West (2024), which readers may find helpful

to see. Our reformulation emphasizes the role of the within-bin structure of the original predictive density and how this structure is preserved in the tilted density. By levering this insight, we are able to derive a novel analytical characterization of the solution for the multiple-histogram case, which is new to the best of our knowledge. In addition, we discuss our computational implementation of the tilting solution in the occasional cases in which the original predictive density from the model does not have mass in some of the target bins from SPF histogram forecasts.

Since relative entropy involes the ratio of the tilted density to the original predictive density, $q(\boldsymbol{X})/p(\boldsymbol{X})$, the solution to the tilting problem is only well-defined if $p(\boldsymbol{X}) > 0$ for all $\boldsymbol{X}$ for which $q(\boldsymbol{X}) > 0$. A necessary condition for a solution to exist is that the support of $p(\boldsymbol{X})$ covers the events for which the SPF provides positive probabilities.

Assumption 1 (Necessary condition: Support of p must cover the SPF histogram events)**.** *For all histogram bins A_i^n for which the SPF provides positive probabilities, the proposal density must also have positive mass. Formally, for all i, n such that $\widehat{Prob}\big(\boldsymbol{X} \in A_i^n\big) > 0$, we require that $p\big(\boldsymbol{X} \in A_i^n\big) > 0$.*

In our application, violations of Assumption 1 are limited to only a few unemployment rate histograms, and quantitatively almost negligible in our application, in that the SPF histogram mass for which the proposal density has zero mass is no larger than 0.2% (and typically zero). A sufficient condition for a solution to exist is that the support of $p(\boldsymbol{X})$ is the entire space of $\boldsymbol{X}$, so that $p(\boldsymbol{X}) > 0$ for all $\boldsymbol{X}$.

Assumption 2 (Highly sufficient condition: Support of p is the entire space of $\boldsymbol{X}$)**.** *The support of the proposal density $p(\boldsymbol{X})$ is the entire space of $\boldsymbol{X}$, so that $p(\boldsymbol{X}) > 0 \; \forall \; \boldsymbol{X}$.*

As argued further below, in Section 3.4, a less restrictive sufficient condition can be formulated for our case of tilting to histograms. For now we assume that Assumption 2 holds, which also covers Assumption 1. In Section 3.7 we discuss an augmented proposal

density that assures the existence of a solution when this sufficient condition (as well as the necessary condition) is not met by the original proposal density.

3.2 Tilting to a single histogram

In the case of tilting to a single histogram, we drop the superscript n to denote the histogram. The entropic tilting problem then amounts to finding a density $q(\boldsymbol{X})$ that matches a given set of marginal distributions, $\{q(A_j)\}_{j=1}^{J}$, and otherwise minimizes the relative entropy $D_x(q \,||\, p)$.[11] The densities $q(\boldsymbol{X})$ and $p(\boldsymbol{X})$ can generally be factorized as

$$q(\boldsymbol{X}) = q(A_j)\, q(\boldsymbol{X}|A_j)\,,\; p(\boldsymbol{X}) = p(A_j)\, p(\boldsymbol{X}|A_j)\,,\; \forall\, j, \boldsymbol{X} \in A_j\,, \tag{8}$$

leading to a corresponding decomposition of relative entropy:

$$D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) = D\left(q(A) \;||\, p(A)\right) + \sum_j q(A_j) \cdot \left\{D\left(q(\boldsymbol{X}|A_j) \;||\, p(\boldsymbol{X}|A_j)\right)\right\},$$

where

$$D\left(q(A) \;||\, p(A)\right) = \sum_j q(A_j) \cdot \log\left(\frac{q(A_j)}{p(A_j)}\right)$$
$$D\left(q(\boldsymbol{X}|A_j) \;||\, p(\boldsymbol{X}|A_j)\right) = \int_{x \in A_j} q(\boldsymbol{X}|A_j) \cdot \log\left(\frac{q(\boldsymbol{X}|A_j)}{p(\boldsymbol{X}|A_j)}\right) dx.$$

Notice that $D\left(q(A) \;||\, p(A)\right)$ is solely determined by the bin probabilities under the baseline density p and the targeted density q. This term is fixed by the choice of bins and proposal density, and it represents the relative entropy between the targeted histogram and its counterpart from the baseline. Crucially, as argued next, the relative entropy between histograms from baseline and the SPF targets is also the *minimized* relative entropy between the baseline density $p(\boldsymbol{X})$ and its tilted counterpart $q(\boldsymbol{X})$. To see this, note that we can

[11] For brevity, we drop the superscript n while discussing the single-histogram case with $n = 1 = N$.

freely choose the *conditionals* of the tilted density $q(\boldsymbol{X}|A)$, to obtain zero relative entropy within each bin. Thus, we can achieve the optimum by simply leaving the densities *inside each bin* unchanged relative to the baseline:

$$D(q(\boldsymbol{X}|A_j) \,||\, p(\boldsymbol{X}|A_j)) = 0 \quad \Leftrightarrow \quad q(\boldsymbol{X}|A_j) = p(\boldsymbol{X}|A_j) \quad \forall\, j. \tag{9}$$

Because relative entropy is nonnegative, this is the minimizing solution to the tilting problem. Intuitively speaking, the minimizing solution reweights draws of $\boldsymbol{X}$ that fall into the same bin A_j with identical weight $q(A_j)/p(A_j)$, and minimizes relative entropy by preserving the shape of the original distribution f inside each bin.

To make the implementation concrete in the single histogram case, suppose we have a simple histogram with just two bins, for inflation next year, with one bin for inflation to exceed 2 percent and the other bin for inflation to be 2 percent or less. Assume that, for the particular year ahead, survey respondents attach equal probabilities of 50 percent to these two bins. Now suppose that, in the baseline forecast density from the BVAR that is represented by Markov Chain Monte Carlo draws from the posterior predictive distribution, the baseline density has 60 percent of forecast draws that exceed 2 percent and 40 percent of draws of inflation of 2 percent. In this case, our entropic tilting solution takes all draws of inflation that are 2 percent or less and reweights (or overweights) them by a factor of 5/4 and takes all draws of inflation exceeding 2 percent and reweights by them by a factor of 5/6. With this solution, the tilted weights integrate to 1, $(5/4 \times 0.4) + (5/6 \times 0.6) = 0.5 + 0.5 = 1$, while satisfying the targeted histogram probabilities of 50 percent in each bin. Moreover, within each bin, the relative weights of the original draws are preserved, so that the shape of the baseline density within each bin is unchanged.

3.3 Analytical solution to tilting with two or more histograms

The multiple-histogram case is a bit more involved than the one-histogram case, because the targeted histograms restrict multiple marginal densities. In practical terms, the histograms tell us the probability of $\boldsymbol{X}$ falling into either set of bins, but nothing about the interdependence between $\boldsymbol{X}$ falling into bins from either histogram.

In the analytics for multiple histograms, we index the histograms by $n = 1, 2, \ldots, N$ and N refers to the number of histograms. Each histogram contains $J(n)$ bins that represent a set of mutually exclusive events. Probabilities of falling into the jth bin of the nth histogram under proposal and targeted histogram are denoted by $p\left(A_j^n\right)$ and $q\left(A_j^n\right)$, respectively.

3.4 The case of two histograms

Considering the case of two histograms, we denote the probabilities pertaining to joint occurrences in bins i and j of the two histograms as follows:[12]

$$p\left(A_i^1, A_j^2\right) \equiv p\left(\boldsymbol{X} \in A_i^1 \cap \boldsymbol{X} \in A_j^2\right) , \tag{10}$$

The SPF prescribes targeted values for, $q\left(A_i^1\right)$ and $q\left(A_j^2\right)$, instead of the joint density $q\left(A_i^1, A_j^2\right)$. To decompose the the relative entropy objective, we factorize the densities:

$$\begin{aligned} q(\boldsymbol{X}) &= q\left(A_i^1, A_j^2\right)\ q\left(\boldsymbol{X}|A_i^1, A_j^2\right) \\ &= q\left(A_i^1\right)\ q\left(A_j^2\right)\ m^q\left(A_i^1, A_j^2\right)\ q\left(\boldsymbol{X}|A_i^1, A_j^2\right) , \qquad (11) \\ p(\boldsymbol{X}) &= p\left(A_i^1, A_j^2\right)\ p\left(\boldsymbol{X}|A_i^1, A_j^2\right) \\ &= p\left(A_i^1\right)\ p\left(A_j^2\right)\ m^p\left(A_i^1, A_j^2\right)\ p\left(\boldsymbol{X}|A_i^1, A_j^2\right) , \qquad (12) \end{aligned}$$

where we have introduced the following "interaction coefficients," that capture the depen-

[12]The definition applies to every $0 < i \leq J(1)$ and $0 < j \leq J(2)$.

dence structure between the two histograms.

$$m^q\left(A_i^1, A_j^2\right) \equiv \frac{q\left(A_i^1, A_j^2\right)}{q\left(A_i^1\right)\, q\left(A_j^2\right)}\,, \qquad m^p\left(A_i^1, A_j^2\right) \equiv \frac{p\left(A_i^1, A_j^2\right)}{p\left(A_i^1\right)\, p\left(A_j^2\right)}\,, \tag{13}$$

These interaction coefficients are measures of association and capture deviations from independence. Formally, the interaction coefficients are also Radon-Nikodym derivatives that describe the change of measure between the product of the marginal distributions $q\left(A_i^1\right)$ and $q\left(A_j^2\right)$ and the joint distribution $q\left(A_i^1, A_j^2\right)$, and the same for $p(\cdot)$.[13] A positive association, $m^q\left(A_i^1, A_j^2\right) > 1$, reflects the case when outcomes tend to occur together (and vice versa for negative associations). From the factorizations in (11) and (12), we obtain a linear decomposition of relative entropy:

$$\begin{aligned} D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) = {} & D\left(q\left(A^1\right) \,||\, p\left(A^1\right)\right) + D\left(q\left(A^2\right) \,||\, p\left(A^2\right)\right) \\ & + \Delta_q\left(m^q\left(A^1, A^2\right) \,||\, m^p\left(A^1, A^2\right)\right) \\ & + E_q\left\{D\left(q\left(\boldsymbol{X}|A^1, A^2\right) \,||\, p\left(\boldsymbol{X}|A^1, A^2\right)\right)\right\}\,, \end{aligned} \tag{14}$$

$$\text{with} \quad \Delta_q\left(m^q\left(A^1, A^2\right) \,||\, m^p\left(A_i^1, A_j^2\right)\right) \equiv \sum_i \sum_j q\left(A_i^1\right)\, q\left(A_j^2\right)\, m^q\left(A_i^1, A_j^2\right)\, \log\left(\frac{m^q\left(A_i^1, A_j^2\right)}{m^p\left(A_i^1, A_j^2\right)}\right), \tag{15}$$

$$\text{and} \quad E_q\left\{D\left(q\left(\boldsymbol{X}|A^1, A^2\right) \,||\, p\left(\boldsymbol{X}|A^1, A^2\right)\right)\right\} \equiv \sum_i \sum_j q\left(A_i^1, A_j^2\right) \cdot D\left(q\left(\boldsymbol{X}|A_i^1, A_j^2\right) \,||\, p\left(\boldsymbol{X}|A_i^1, A_j^2\right)\right)\,. \tag{16}$$

As in the one-histogram case, the first term in (14) is fixed by the choice of bins and

[13] The product of the marginal distributions characterizes a (hypothetical) joint distribution that would arise if events in both histograms were independent. Note that $m^q\left(A_i^1, A_j^2\right) = 1$ if and only if events in both histograms are independent under $q(\cdot)$.

the targeted histogram probabilities; we refer to this term as "marginal entropy," since it captures the relative entropy between the marginal histogram densities. The last term in (14) captures an average of the relative entropies between the conditional densities inside each of the joint cross-histogram events, as defined in (16), which are not directly restricted by the histograms. As before, the optimal choice is to leave the within-cross-event densities unchanged relative to the baseline.

Proposition 1 (Optimal tilting leaves within-cross-event densities unchanged)**.** *The optimal choice of* $q(\cdot)$ *to minimize the relative entropy objective in* (14) *while satisfying the constraints in* (**??**) *is such that the relative entropy between the within-cross-event densities is zero. The result extends directly also to the case of more than two histograms,* $N \geq 2$*, so that* $q\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right) = p\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right) \forall\, i, j, \ldots$

Proof. The result follows from the ability to choose $q\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right)$ without constraints and separately from $q\left(A_i^1, A_j^2, \ldots\right)$, where our notation anticipates the generalization to more than two histograms. In addition, the non-negativity of relative entropy, $D\left(q \,||\, p\right)$ implies that its minimum is achieved by setting the within-cross-event relative entropies to zero, which is achieved if and only if $q = p$ (Cover and Thomas, 2006). In the case of within-cross-event relative entropies, $D\left(q\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right) \,||\, p\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right)\right)$, this translates to $q\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right) = p\left(\boldsymbol{X}|A_i^1, A_j^2, \ldots\right) \forall\, i, j, \ldots$ □

As a corollary of Proposition 1, we also obtain a sufficient condition for the existence of a solution to the tilting problem that is less restrictive than Assumption 2.

Assumption 3 (Sufficient condition: Support of p covers all cross-histogram events)**.** *The proposal density has positive mass for all joint events* $A_i^1 \cap A_j^2 \cap \ldots$ *Formally, we require that* $p\big(\boldsymbol{X} \in A_i^1 \cap A_j^2 \cap \ldots\big) > 0$ *for all* $i, j, \ldots$

Proposition 1 leaves the second term in (14), $\Delta_q\left(m^q\left(A^1, A^2\right) \,||\, m^p\left(A^1, A^2\right)\right)$, as the only remaining component of relative entropy to be affected by the optimal choice of $q(\boldsymbol{X})$. We refer to $\Delta_q\left(\cdot||\cdot\right)$ as "interaction entropy," and it measures the entropic cost of

changing the interaction (or association) structure between the two histograms. We need to choose the joint bin probabilities $q\left(A_i^1, A_j^2\right)$, or equivalently the interaction coefficients $m^q\left(A_i^1, A_j^2\right)$, to minimize interaction entropy while satisfying the constraints imposed by the targeted marginal histogram probabilities:[14]

$$\min_{m^q(A^1,A^2)} \Delta_q\left(m^q\left(A^1, A^2\right) \,||\, m^p\left(A^1, A^2\right)\right) \tag{17}$$

$$\text{subject to} \quad \sum_i q\left(A_i^1\right)\, m^q\left(A_i^1, A_j^2\right) = \sum_i q\left(A_i^1 | A_j^2\right) = 1\,, \qquad \forall\, j\,, \tag{18}$$

$$\text{and} \quad \sum_j q\left(A_j^2\right)\, m^q\left(A_i^1, A_j^2\right) = \sum_j q\left(A_j^2 | A_i^1\right) = 1\,, \qquad \forall\, i\,, \tag{19}$$

The first order conditions for minimizing interaction entropy imply the following:

$$m^q\left(A_i^1, A_j^2\right) \quad = \quad m^p\left(A_i^1, A_j^2\right)\, \cdot \phi_i^1 \cdot \phi_j^2 \qquad\qquad \forall i, j \tag{20}$$

where ϕ_i^1 and ϕ_j^2 are Lagrange multipliers, each for one of the bin-specific constraints in (18) and (19). It is convenient to rescale the Lagrange multipliers by the corresponding ratio of bin-event probabilities, and we define

$$\psi_i^n \equiv \frac{q\left(A_i^n\right)}{p\left(A_i^n\right)} \cdot \phi_i^n \tag{21}$$

At the optimum, the rescaled Lagrange multipliers, ψ_i^1 and ψ_j^2, must satisfy the following first-order conditions:

$$\psi_i^1 \cdot \left(\sum_j \psi_j^2 \cdot p\left(A_i^1, A_j^2\right)\right) = q\left(A_i^1\right)\,, \qquad \forall\; i \tag{22}$$

$$\text{and} \quad \psi_j^2 \cdot \left(\sum_i \psi_i^1 \cdot p\left(A_i^1, A_j^2\right)\right) = q\left(A_j^2\right)\,, \qquad \forall\; j \tag{23}$$

[14]Note that, as a consequence of (18) and (19), we have $\sum_j q\left(A_j^2 | A_i^1\right) q\left(A_i^1\right) = q\left(A_j^2\right)$ and $\sum_i q\left(A_i^1 | A_j^2\right) q\left(A_j^2\right) = q\left(A_i^1\right)$. Moreover, $\sum_j \sum_i q\left(A_i^1, A_j^2\right) = 1$ also holds, since the targeted (marginal) histogram probabilities sum to one.

Without convenient closed-form solutions for ψ_i^1 and ψ_j^2 available, we obtain the solution by iterating over (22) and (23). This procedure is an instance of "iterative proportional fitting" (IPF) which has a longstanding tradition in the statistics literature.

Algorithm 1 (Iterative proportional fitting (for $N = 2$))**.** *This is an iterative algorithm with iteration steps indexed by s. For $s = 0$ set initial values $\psi_i^1\{s\} = q\left(A_i^1\right) \big/ p\left(A_i^1\right)$ and $\psi_j^2\{s\} = q\left(A_j^2\right) \big/ p\left(A_j^2\right)$. Iterate over the following until convergence:*

1. *Given $\psi_j^2\{s-1\}\ \forall\ j$, for each event i in A^1, set $\psi_i^1\{s\}$ to satisfy*

$$\psi_i^1\{s\} \cdot \left(\sum_j \psi_j^2\{s-1\} \cdot p\left(A_i^1, A_j^2\right) \right) = q\left(A_i^1\right),\ \forall\ i\,. \tag{24}$$

2. *Given $\psi_i^1\{s\}\ \forall\ i$, for each event j in A^2, set $\psi_j^2\{s\}$ to satisfy*

$$\psi_j^2\{s\} \cdot \left(\sum_i \psi_i^1\{s\} \cdot p\left(A_i^1, A_j^2\right) \right) = q\left(A_j^2\right),\ \forall\ j\,. \tag{25}$$

Note that each step enforces the corresponding constraints of (18) *and* (19)*, so that all constraints are assured to hold when the solution has converged at $\psi_i^1\{s\} = \psi_i^1\{s-1\}$ and $\psi_j^2\{s\} = \psi_j^2\{s-1\}\ \forall\ i, j$.*

Convergence of Algorithm 1 is assured by the general properties of the entropic tilting problem discussed by Robertson, Tallman, and Whiteman (2005) as well as Tallman and West (2022) and West (2024). In our experiences, it also converges quite quickly, often in twenty or less steps with each step just taking a few seconds to compute (see Figure 15 in the appendix for details).

The ratio of tilted to untilted probabilities, also known as "tilting weights", can conve-

niently be expressed as simple cross-product of (scaled) Lagrange multipliers:[15]

$$w_{i,j} \equiv \frac{q\left(A_i^1, A_j^2\right)}{p\left(A_i^1, A_j^2\right)} = \psi_i^1 \cdot \psi_j^2 \qquad \text{with} \quad \psi_i^n \equiv \frac{q\left(A_i^n\right)}{p\left(A_i^n\right)} \cdot \phi_i^n \tag{26}$$

The expression for the optimal tilting weights in (26) is very useful for computational efficiency, since the $J(1) \times J(2)$ values for the weights $w_{i,j}$ are pinned down by only $J(1) + J(2)$ Lagrange multipliers ψ_i^1 and ψ_j^2. For sake of illustration, consider two histograms, each with $J(1) = J(2) = 10$ bins, and the $10 \times 10 = 100$ optimal tilting weights $w_{i,j}$ are characterized by cross-products of merely $10 + 10 = 20$ Lagrange multipliers, ψ_i^n.

As argued below, the analytics presented here for $N = 2$ straightforwardly extend to $N \geq 2$, so that the computational benefits emanating from (26) become even more relevant. For ease of notation, suppose each histogram has the same number of bins, J. In this case, the number of Lagrange multipliers is linear in N (namely $N \times J$), whereas the number of tilting weights grows exponentially, by J^N. More generally, with $J(n)$ for the nth histogram, the number of tilting weights grows by $\prod_{n=1}^{N} J(n)$, while the number of Lagrange multipliers grows by $\sum_{n=1}^{N} J(n)$.

3.5 The general case of three and more histograms

Our approach easily generalizes to case of three and more histograms. As in the two-histogram case, relative entropy can be decomposed into (a) the sum of relative entropies for the marginal histograms, (b) relative entropy of the cross-product densities ("interaction entropy"), and (c) the relative entropy of the within-cross-event densities. As before, the marginal entropy is fixed by the problem, and the within-cross-event entropies can always be set to zero, so that only the minimization of interaction entropy remains.

[15] Henceforth we will be referring to ψ_i^n simply as Lagrange multipliers.

With $N = 3$, the optimal solution is characterized by the following:

$$\frac{q\left(A_i^1, A_j^2, A_k^3\right)}{p\left(A_i^1, A_j^2, A_k^3\right)} = \psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3, \tag{27}$$

where ψ_i^1, ψ_j^2, ψ_k^3 are defined analogously to (26).

The optimal Lagrange multipliers satisfy the following set of conditions:

$$\psi_i^1 \cdot \left(\sum_j \sum_k p\left(A_i^1, A_j^2, A_k^3\right) \cdot \psi_j^2 \cdot \psi_k^3 \right) = q\left(A_i^1\right) , \qquad \forall\, i\,, \tag{28}$$

$$\psi_j^2 \cdot \left(\sum_i \sum_k p\left(A_i^1, A_j^2, A_k^3\right) \cdot \psi_i^1 \cdot \psi_k^3 \right) = q\left(A_j^2\right) , \qquad \forall\, j\,, \tag{29}$$

$$\psi_k^3 \cdot \left(\sum_i \sum_j p\left(A_i^1, A_j^2, A_k^3\right) \cdot \psi_i^1 \cdot \psi_j^2 \right) = q\left(A_k^3\right) , \qquad \forall\, k\,, \tag{30}$$

which are a direct extension of the two-histogram case. Together with the iterative proportional fitting scheme described in Algorithm 1, the first-order conditions are straightforward to generalize to arbitrary N. The cross-product structure of the optimal tilting weights in (27) also affords a convenient characterization of the relative entropy between q and p at the optimum, which is expressed in the following proposition.

Proposition 2 (Relative entropy of optimizing solution)**.** *The optimal value of relative entropy in the entropic tilting problem of* (7) *has the following linear decomposition:*

$$D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) = \sum_{n=1}^{N} \sum_{i=1}^{J(n)} q\left(A_i^n\right) \cdot \log\left(\psi_i^n\right) \tag{31}$$

$$= \sum_n \sum_i q\left(A_i^n\right) \cdot \log\left(\frac{q\left(A_i^n\right)}{p\left(A_i^n\right)}\right) + \sum_n \sum_i q\left(A_i^n\right) \cdot \log \phi_i^n \tag{32}$$

where the first term in (32) *corresponds to the sum of relative entropies between the marginal distributions for each histogram, and the second term captures the adjustments in tilting weights necessary to preserve the correlation structure of the proposal density* $p(\cdot)$ *as much as possible.*

Proof. The decomposition of relative entropy in the entropic tilting problem follows from the properties of relative entropy divergence and the structure of the constraints. For ease of exposition, we derive this for the case of three histograms. By Proposition 1, the optimizing solution leaves the within-cross-histogram event densities unchanged. The relative entropy at the optimum then reduces to

$$D\left(q(\boldsymbol{X}) \,||\, p(\boldsymbol{X})\right) = D\left(q(A_i^1, A_j^2, A_k^3) \,||\, p(A_i^1, A_j^2, A_k^3)\right) \tag{33}$$

$$= \sum_i \sum_j \sum_k q\left(A_i^1, A_j^2, A_k^3\right) \log\left(w_{i,j,k}\right) \tag{34}$$

$$= \sum_i \sum_j \sum_k q\left(A_i^1, A_j^2, A_k^3\right) \log\left(\psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3\right) \tag{35}$$

and the result follows from $\log\left(\psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3\right) = \log\left(\psi_i^1\right) + \log\left(\psi_j^2\right) + \log\left(\psi_k^3\right)$ and $\sum_j \sum_k q\left(A_i^1, A_j^2, A_k^3\right) = q(A_i^1)$ etc. □

3.6 Digression: The case of independent distributions

Before turning to the computational implementation of the tilting solution, we briefly discuss the special case in which the original density $p(\boldsymbol{X})$ generates draws that are independent across histograms. Relatedly, we also describe a feasible choice of tilting weights that targets the product density, $q^{iid}\left(A_i^1, A_j^2, A_k^3\right) = q\left(A_i^1\right) \cdot q\left(A_j^2\right) \cdot q\left(A_k^3\right)$, which would be the optimal solution if the proposal density saw the histograms as independent from one another. For sake of exposition, we consider the case of three histograms with indices i, j, and k, as in the previous section, but the discussion straightforwardly extends to arbitrary N.

Lemma 1 (Unit interaction entropy in case of independence)**.** *When the original density, $p\left(A_i^1, A_j^2, A_k^3\right)$, is identical to the product density of its marginals, the interaction*

coefficient equals one:

$$p\left(A_i^1, A_j^2, A_k^3\right) = p\left(A_i^1\right) \cdot p\left(A_j^2\right) \cdot p\left(A_k^3\right) \Leftrightarrow m^p\left(A_i^1, A_j^2, A_k^3\right) = 1\,, \quad \forall\, i, j, k$$

Proof. The result follows directly from the definition of the interaction coefficients as the ratio between the joint density and the corresponding product of its marginals, see (13) for the case of $N = 2$. □

Lemma 2 (Independence tilt)**.** *A feasible—though not necessarily entropy-minimizing—choice of tilting weights targets the product density,* $q^{iid}\left(A_i^1, A_j^2, A_k^3\right) = q\left(A_i^1\right) \cdot q\left(A_j^2\right) \cdot q\left(A_k^3\right)$*. The product density is straightforward to implement with the following tilting weights:*

$$w_{i,j,k}^{iid} \equiv \frac{q\left(A_i^1\right) \cdot q\left(A_j^2\right) \cdot q\left(A_k^3\right)}{p\left(A_i^1, A_j^2, A_k^3\right)} \tag{36}$$

and it is straightforward to verify that all marginal bin probabilities satisfy the targeted (marginal) densities, $q\left(A_i^n\right)\ \forall\, i, n$*.*

We refer to this choice of tilting weights described in (36) as $w_{i,j,k}^{iid}$ since, as argued next, it would be the optimal choice if the proposal density saw the histograms as independent from one another.

Proposition 3 (Optimal tilting when histograms are independent under the proposal)**.** *When the histograms are independent under the proposal, the optimal tilting weights are given by* $w_{i,j,k}^{iid}$ *in* (36)*, which target the product density* $q^{iid}\left(A_i^1, A_j^2, A_k^3\right) = q\left(A_i^1\right) \cdot q\left(A_j^2\right) \cdot q\left(A_k^3\right)$*. Equivalently, the entropy-minimizing (unscaled) Lagrange multipliers are all identical to one in this case,* $\phi_i^1 = \phi_j^2 = \phi_k^3 = 1$*.*

Proof. When the histograms are independent under the proposal, we have $p\left(A_i^1, A_j^2, A_k^3\right) =$

$p\left(A_i^1\right) \cdot p\left(A_j^2\right) \cdot p\left(A_k^3\right)$, so that

$$p\left(A_i^1, A_j^2, A_k^3\right) = p\left(A_i^1\right) \cdot p\left(A_j^2\right) \cdot p\left(A_k^3\right) , \tag{37}$$

$$\implies \quad w_{i,j,k}^{iid} = \frac{q\left(A_i^1\right)}{p\left(A_i^1\right)} \cdot \frac{q\left(A_j^2\right)}{p\left(A_j^2\right)} \cdot \frac{q\left(A_k^3\right)}{p\left(A_k^3\right)} , \tag{38}$$

which corresponds to setting the (unscaled) Lagrange-multipliers all to one, $\phi_i^n = 1$, so that $\psi_i^n = q(A_i^n)/p(A_i^n)$, $\forall\ i, n$. Moreover, since $m^p\left(A_i^1, A_j^2, A_k^3\right) = 1$, the interaction entropy reduces to the relative entropy between $q\left(A_i^1, A_j^2, A_k^3\right)$ and the product density:

$$\Delta_q\left(m^q\left(A^1, A^2, A^3\right) \,||\, 1\right) = D\left(\, q\left(A^1, A^2, A^3\right) \,||\, q\left(A^1\right) \cdot q\left(A^2\right) \cdot q\left(A^3\right)\right) , \tag{39}$$

which is set to zero by letting $q\left(A_i^1, A_j^2, A_k^3\right) = q\left(A_i^1\right) \cdot q\left(A_j^2\right) \cdot q\left(A_k^3\right)$. Since relative entropy cannot be negative, it is also the minimizing choice. □

An immediate upshot of Proposition 3 is that sequentially targeting individual histograms is (jointly) optimal when the proposal distribution sees the histogram events as mutually independent. In this case, the entropy-minimizing density preserves the proposed independence between the histograms and implements the product density $q^{iid}(\cdot)$ discussed in Lemma 2.

3.7 Computational challenges and implementation

By virtue of Proposition 1, the tilting weights for the entropic tilting problem with histogram constraints are identical for all outcomes within a given cross-histogram event. Moreover, while these weights depend on the cross-histogram event probabilities of the propsal, $p(\boldsymbol{c})$, they do not depend on the within-bin densities, $p(X|\boldsymbol{c})$. As such, the entropic tilting problem with histogram constraints has a compact structure.

However, in practice, entropic tilting to histograms faces also two main challenges. First, the necessary and sufficient conditions for the existence of a solution to the tilting

problem may not be satisfied by the proposal density. Second, when the sufficient conditions are met, the computational burden of solving for the optimal tilting weights can be substantial, especially as the number of histograms increases.

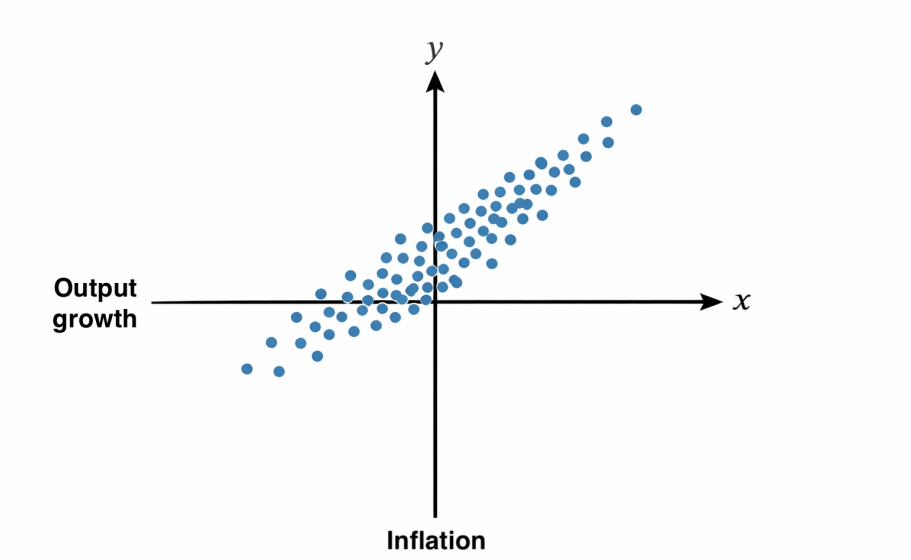
This section develops an approach to address both of these challenges. We begin by describing the relevance of the sufficient condition in Assumption 3, and how a solution may exist even when the condition is not fully satisfied. Cases where Assumption 3 is not met reflect sparsity in the proposal density, which in itself offers a potential avenue to alleviate the computational burden of solving for the optimal tilting weights. We then describe our solution to combine the benefits of sparsity within an augmented proposal density that ensures the sufficient conditions for the existence of a solution to the tilting problem are met.

3.7.1 Sparse proposals and failure to meet the sufficient condition

To appreciate the conceptual relevance of the sufficient condition in Assumption 3 we consider the following schematic example. Figure 1 displays a setting with two histograms, one for inflation and one for output growth, that have two bins each. The proposal density provides draws for high and low inflation, as well as high and low output growth, however it lacks joint occurrences of low growth and high inflation. In this configuration, the necessary condition in Assumption 1 is satisfied, since the proposal density has mass in all bins of the individual histograms. However, the sufficient condition in Assumption 3 is not satisfied, since the proposal density does not have mass in the joint event of low growth and high inflation. Having zero mass for cross-histogram events is an instance of sparsity in the proposal density in $p(\cdot)$.

Since the SPF puts no explicit target on the combination of high inflation and low growth, tilting to both histograms jointly can still work as long as the targeted (marginal) probabilities for the events whose joint occurrence is not supported by the proposal density, are not too extreme. Specifically, as detailed in Appendix A, a solution to the tilting

Figure 1: Illustration of case with zero mass on some joint cross-histogram events



Notes: The illustration shows a two-dimensional scatter plot of draws from a hypothetical proposal density for GDP growth and inflation, with outcomes for each sorted in two bins. Blue dots represent draws from the proposal density. The plot demonstrates a case where the proposal density has support in three of the four quadrants (high growth with low inflation, high growth with high inflation, and low growth with low inflation) but lacks any draws in the lower-left quadrant (low growth with high inflation), illustrating a violation of the sufficient condition in Assumption 3.

problem satisfying all constraints may or may not be feasible, depending on whether the the sum of the targeted probabilities for low growth and high inflation is below or above 100%. Similar reasoning (albeit with more complex conditions) applies to cases of more than two histograms and with more than two bins.

In principle, provided a solution to the tilting problem exists, sparsity in $p(\cdot)$ can be computationally very appealing.[16] In fact, sparsity becomes a computational necessity as the number of histograms and bins per histograms grows. For example, in our application we have $N = 10$ histograms having $J = 10$ bins each (some have 11), leading to over 10^{10}, i.e. 10 billion, cross-histogram events. Such a larger number of events $|\mathcal{C}|$ poses significant challenges for processing the (discrete) densities $p(\boldsymbol{c})$ and $q(\boldsymbol{c})$. Most notably, the memory

[16]Of course, for a solution to exist even though Assumption 3 fails to be met, the much milder necessary condition of Assumption 1 must be satisfied, as is the case in the example of Figure 1.

requirement of an array of numbers of size 10^{10} is about 80 GB, and simultaneously processing $p(\boldsymbol{c})$ and $q(\boldsymbol{c})$ then requires twice that amount in RAM capacity, not to mention execution times. By contrast, in the illustrative example of Figure 1, one quarter of its elements of $p(\cdot)$ are zero, and in practice we find that sparsity can extend to more than nine tenth of $p(\boldsymbol{c})$, allowing for massively reduced storage needs and improved computation times. Sparse arrays not only afford compressed storage, but also reduced execution times since additive expressions of elements equal to zero can be omitted, and the result of multiplicative expressions involving zeros are also known. Computing languages specialized in matrix operations, such as MATLAB or Julia, provide such fast and efficient methods for sparse arrays, and our implementation is based on MATLAB.

In fact, sparsity in $p(\cdot)$ is a natural implication of correlation (or more generally, interdependence) of outcomes targeted by different histograms. Such interdependence can reflect persistence across different forecast horizons for the same economic variables, such as inflation this year and next year, or the macroeconomic relationship between different economic variables, such as unemployment and GDP growth, or a combination of these factors. Of course, many models, such as BVARs with (conditionally) Gaussian shocks, have in principle continuous support for every outcome on the real line. However, simulated-based density approximations will fail to deliver draws for highly unlikely outcomes even when using a large (but finite) number of draws. Such unlikely constellations arise, for example, when considering jumps from the top bin of the histogram for a given variable in one year to the bottom bin in the next year, followed by another jump to the top bin in the year thereafter paired with another sequence of moves for further variables covered by other histograms. That said, when outcomes are independent and unbounded, the proposal should deliver positive mass for every cross-histogram event $\boldsymbol{c}$ and sparsity resulting from the finiteness of the Monte Carlo sample would be undesirable. However, Proposition 3 enables us to use sequential computations for the case of a proposal that sees cross-histogram events as independent from one another, without need for processing

all elements of $p()$ jointly. This insight motivates use of an augmented proposal, mixing properties of a sparse model density with an independent uniform component applied to all cross-histogram events $\boldsymbol{c}$, as described in the remainder of this section.

3.7.2 Augmented proposal density to ensure sufficient conditions are met

In order to handle cases where Assumption 3 (or the necessary condition in Assumption 1) is not met, we use an augmented proposal density that has mass in all joint cross-histogram events. The augmented density is a mixture of the original proposal, $p^0(\cdot)$, and a uniform density for the occurrence of each cross-histogram event.

$$p(\boldsymbol{c}) = \big(1-\omega\big)\, p^0(\boldsymbol{c}) + \omega \cdot \frac{1}{|\mathcal{C}|}\,, \qquad\qquad \boldsymbol{c} \equiv A_i^1 \cap A_j^2 \cap A_k^3 \cap \ldots \tag{40}$$

where $p^0(\cdot)$ denotes the original proposal density (say, from the BVAR), and $|\mathcal{C}|$ is the number of elements in the set of cross-histogram events $\mathcal{C}$ with typical element $\boldsymbol{c}$.[17] All the analytics presented in the preceding sections are understood to be applied to the augmented density $p(\cdot)$ rather than $p^0(\cdot)$.

For the mixture weight ω, we choose a small value, so that the augmented proposal is close to the original proposal, placing only minimal weight on the events for which the original proposal density does not generate draws.[18] Nevertheless, even for a small ω the augmented density satisfies Assumptions 1 and 3 by placing positive mass on all cross-histogram events. As discussed further below, the fact that the uniform density is independent across cross-histogram events, makes cases of large $|\mathcal{C}|$ computationally

[17] Formally, the set of all possible cross-histogram events is described by the set of possible tuples for their bin indices:

$$\mathcal{C} = \left\{ \bigcap_{n=1}^{N} A_{c_n}^n \text{ with } \boldsymbol{c} = (c_1, \ldots, c_N) \in \mathbb{N}^N \;:\; 1 \le c_n \le J(n) \text{ for all } n = 1, \ldots, N \right\}$$

where $J(n)$ denotes the number of bins in the nth histogram, and the total number of joint cross-histogram events becomes $|\mathcal{C}| = \prod_n J(n)$.

[18] In our application, we set $\omega = 1e^{-4}$. With $|\mathcal{C}| \approx 10^{10}$ that leaves $\omega/|\mathcal{C}| \approx 10^{-14}$ and thus just above a level of machine accuracy of a little under 10^{-15}.

tractable, since it enables us to lever the insights of Proposition 3 to compute the uniform component of the tilting weights separately for each histogram and independently of one another.

The uniform component in the mixture of (40), leaves the within-joint-event density, $p(X|\boldsymbol{c})$ unspecified. We maintain the original proposal density for cross-histogram events $\boldsymbol{c}$, for which p^0 has mass. Formally, for all $\boldsymbol{c}$ where $p^0(\boldsymbol{c}) > 0$, we set $p(X|\boldsymbol{c}) = p^0(X|\boldsymbol{c})$ and use the draws generated by p^0 for the cross-histogram event $\boldsymbol{c}$. The augmented proposal is, of course, motivated by the occurrence of cases where the original proposal density does not have mass, $p^0(\boldsymbol{c}) = 0$. For these cases, we generated within-bin values by projections of $\boldsymbol{X_t}$ onto the center of the bins involved in the cross-histogram event, as detailed in an appendix.

While we find use of the augmented density necessary at many forecast origins (and various tilting targets), the amount of mass ultimately assigned to cross-histogram events for which the original proposal density does not generate draws, is very small. As shown in Figure 16 in the appendix, apart from the onset of COVID-19, the mass assigned to zero-mass events rarely exceeds 10%, and the large amounts of mass (of over 60%) assigned in early 2020 is driven by the need to account for unemployment rate histograms.

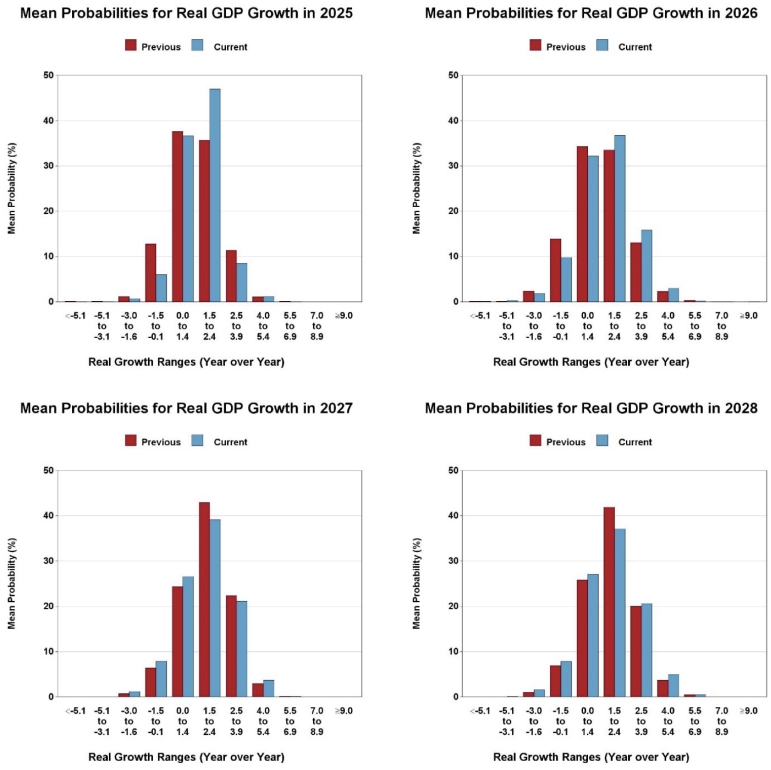
4 Data and model specification

This section details the SPF forecasts used in entropic tilting, the real-time data used for model estimation and forecasting, and the forecasting model.

4.1 Survey data from the SPF

Our analysis uses forecasts of real GDP growth, the unemployment rate, and core PCE inflation from the U.S. Survey of Professional Forecasters (SPF). The SPF collects forecasts quarterly, and they are publicly available on the Federal Reserve Bank of Philadelphia's

Figure 2: SPF histogram for GDP growth per 2025Q3

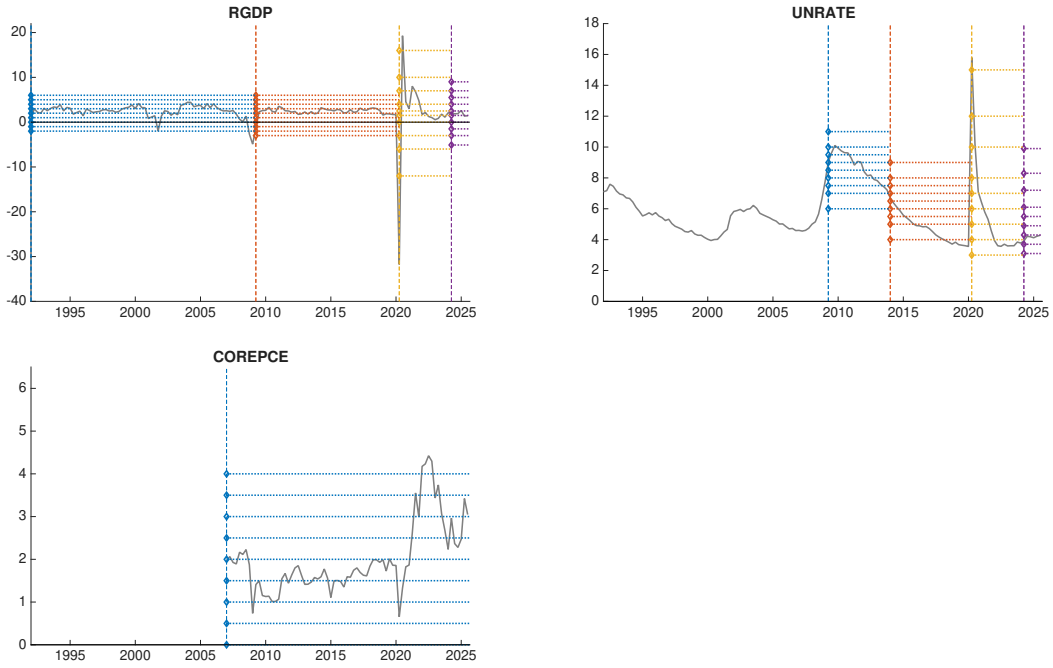


website.[19] In particular, we consider the SPF's density forecasts, which are collected in the form of probability histograms for fixed-event, calendar-year forecasts, for (variously) the current year up through three years ahead. These forecasts refer to percent changes in annual average levels of GDP; the annual average level of the unemployment rate; and the Q4/Q4 percent change in the core PCE price index. Figure 2 shows the histogram forecasts for GDP growth published for 2025Q3. The histograms are constructed from individual forecaster responses, which are aggregated into bins. While our BVAR specification and forecast evaluation will include growth in non-residential fixed investment, growth in payroll employment, and the 5-year Treasury yield, the SPF does not collect or publish density forecasts of these variables.

The availability and specifics of histogram forecasts have varied over time. Since 1981, the SPF has always included current- and next-year density forecasts of GDP growth.

[19]The Federal Reserve Bank of Philadelphia took over management of the SPF in 1992.

Figure 3: SPF histogram bins over time (since 1992)



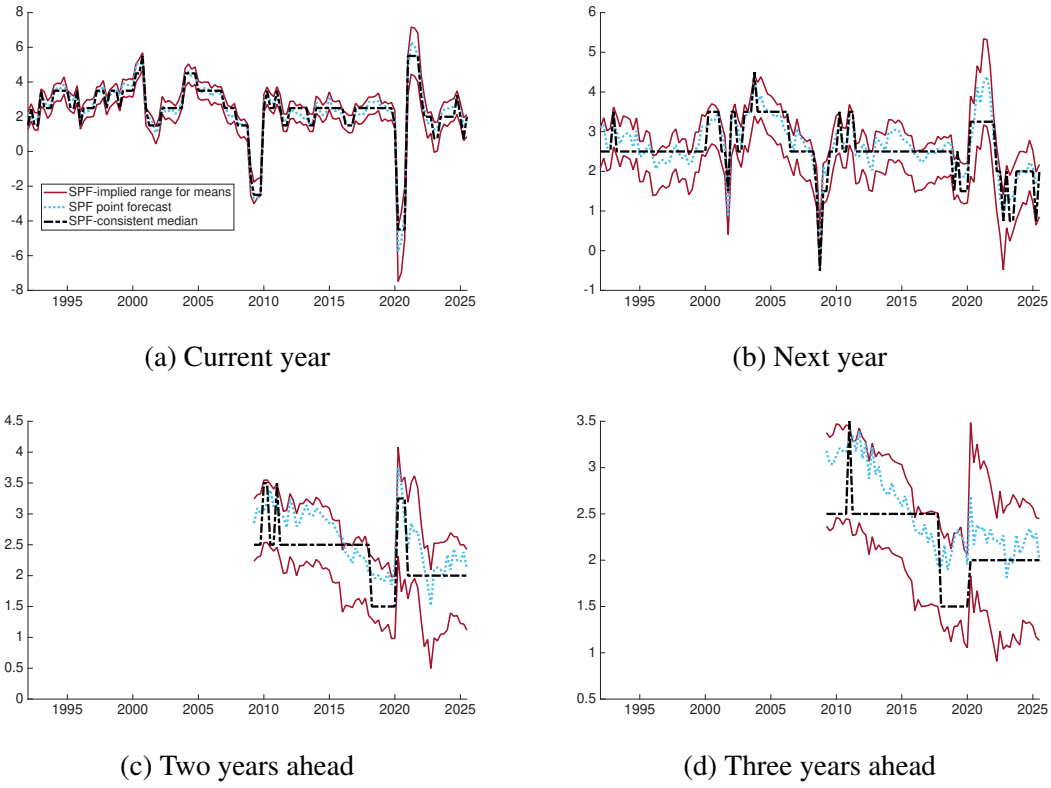
Note: Solid grey lines provide historical data on growth and unemployment (current vintage).

In 2007Q1, the survey added current- and year-ahead histogram forecasts of core PCE inflation, and has maintained this coverage through the present. In 2009Q2, the survey extended the GDP horizons to include 2- and 3-year-ahead histograms and added histogram forecasts for the unemployment rate for the current year through 3 years ahead. Reflecting the availability of real-time data on core PCE inflation (more below), the earliest histogram forecasts we use are for 1996Q1.

The ranges underlying the histogram forecasts (in most cases) have also changed over time, leading to variation in the bins and their widths. Figure 3 shows time series of the bins for the histograms of real GDP growth, unemployment rate, and core PCE inflation forecasts used by the SPF. For example, the SPF lowered the histogram ranges for the unemployment a few years after the Great Recession ended, widened them sharply with the COVID-19 pandemic, and narrowed the unemployment ranges a few years later.

Throughout, we collect probabilistic forecast for the average SPF forecaster ("consensus histogram"), which is consistent with treating the ensemble of individual forecasts as

Figure 4: SPF-implied mean and median forecasts for real GDP growth



(a) Current year

(b) Next year

(c) Two years ahead

(d) Three years ahead

Note: SPF-consistent median is the midpoint of the bin containing the median. Upper and lower bounds for SPF-consistent means are computed from lower and upper bounds of the bins. The point forecasts are the average of individual forecasters' point forecasts. Data is shown for available SPF histograms since 1992.

an equally weighted linear prediction pool. Equally weighted prediction pools have been shown to often lead to good forecast performance (for example, see Clark and Mertens (2024) for a recent survey).[20] Moreover, as argued next, the SPF consensus histograms display good coherence with SPF point forecasts. By contrast, Engelberg, Manski, and Williams (2009) and Clements (2025, 2010) have documented notable inconsistencies between point and histogram predictions of individual SPF participants. For the case of GDP growth forecasts, Figure 4 compares SPF point forecasts against potential mean and median values consistent with the SPF histograms. Following Engelberg, Manski, and Williams (2009) and Clements (2010), a range of SPF-consistent means is computed while

[20]Studies such as Chernis, et al. (2025) and Diebold, Shin, and Zhang (2023) (and other references therein) have developed alternative pooling schemes for individual SPF histogram forecasts.

taking the lower and upper bounds of each bin as the location of the probability mass within that bin. As can be seen, the SPF point forecasts lie within or very close to the range of SPF-consistent means for all forecast horizons and for most of the times.[21]

For sake of comparison, we also compute SPF-consistent medians from the midpoints of the bins containing the median, which is found by accumulating bin probabilities until reaching 50%. (Of course, the true medians are not known and could lie within the range of upper and lower limits of the respective median bins which we omit for sake of brevity). Not surprisingly, the SPF-consistent medians are less sensitive to quarterly changes in the outlook, in particular at longer forecast horizons, which reflects inertia in the identity of the median bin, as well as the infrequent nature of changes in bin limits. That said, SPF-consistent medians broadly track the point-forecast trajectories and often lie within the range of histogram-consistent mean forecasts.[22]

Overall, this coherence between point and histogram forecasts for the consensus SPF histograms suggests that tilting to histogram information will not produce results that are wildly at odds with point forecast information. Of course, point forecast information from the SPF has a well-established track record of predictive value. For the remainder of this paper, we will ignore information on point forecasts, which also allows us to sidestep the question as to whether those should be interpreted as means or modal forecasts or other metrics. Instead, we solely focus on the predictive value contained in histograms.

4.2 Real-time data for model estimation and forecast evaluation

In keeping with common macroeconomic forecasting, we obtain baseline forecasts from models of major macroeconomic aggregates, at the quarterly frequency: real GDP growth, growth in non-residential fixed investment, growth in payroll employment, the unemploy-

[21] As shown in Figures 13 and 14 in the appendix, similar results apply also to the cases of of SPF histograms for core inflation and the unemployment rate.

[22] Nevertheless, the SPF-consistent median values for GDP growth typically run below the actual point forecasts, suggestive of skew in the growth distribution (when point forecasts are interpreted as means).

ment rate, core PCE inflation, and the 5-year Treasury yield.[23] Real-time data vintages for GDP, investment, employment, the unemployment rate, and the core PCE price index are taken from the Federal Reserve Bank of Philadelphia's Real-Time Data Set for Macroeconomists (RTDSM).[24] We obtained the 5-year Treasury yield from the FRED database of the Federal Reserve Bank of St. Louis.

Our analysis of out-of-sample forecasts uses real-time data vintages from 1996Q1 (the first for which core PCE data are available) through 2025Q3. 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 each forecast origin t starting with 1996:Q1, we use the real-time data vintage t containing data through $t-1$ to estimate the forecast models and construct forecasts for periods t and beyond. To evaluate the accuracy of the real-time forecasts, we use the second available (in the quarterly vintages of the RTDSM) estimates of the real-time measured variables as actuals in evaluating forecast accuracy. Our models are estimated using data starting in 1959.

4.3 BVAR-SVO-t model

We produce baseline macroeconomic forecasts using the BVAR-SVO-t model developed in Carriero, et al. (2024), but modified so that stochastic volatility follows an AR(1) rather than random walk process. The BVAR takes the form::

$$y_t = \Pi_0 + \Pi(L)y_{t-1} + v_t\,, \qquad\qquad v_t \sim N(0, \Sigma_t), \tag{41}$$

where y_t is a vector of N observables, $\Pi(L) = \sum_{i=1}^{p} \Pi_i \; L^{i-1}$ is a pth-order lag polynomial of VAR coefficients, and v_t denotes the VAR's residuals. We denote the vector of stacked

[23] Quarterly levels of the unemployment rate, payroll employment, and Treasury yield are monthly averages. Growth of GDP, investment, and employment and PCE inflation are constructed as 400 times log differences of the quarterly levels of the variables.

[24] In the case of employment, for which the RTDSM makes monthly vintages available, in keeping with RTDSM timing conventions we assigned the vintage of the middle month of each quarter as the associated quarterly vintage.

coefficients contained in $\{\Pi_i\}_{i=0}^{p}$ as Π. All models are specified with non-conjugate priors for Π and Σ_t. In the results presented below, the innovation variance matrix Σ_t varies over time. The online appendix shows that a homoskedastic (constant Σ) BVAR yields very similar impacts of tilting to SPF histograms.

The innovation vector in the model consists of a vector of Gaussian innovations scaled by three time-varying variance components, for: (i) stochastic volatility; (ii) rare, transitory, and large — i.e., outlier — changes in volatility; and (iii) transitory changes in volatility that are more frequent but less extreme in impact (consistent with draws from the tails of a fat-tailed distribution). With this stochastic structure, the VAR's innovation v_t can be seen as mixed Gaussian instead of Gaussian. More specifically, the innovation vector and its covariance matrix take the forms:

$$v_t = A^{-1}\,\Lambda_t^{0.5}\,O_t\,Q_t\,\varepsilon_t,\;\; \varepsilon_t \sim N(0,I),\;\; \Sigma_t = A^{-1}\,O_t\,Q_t\Lambda_t\,Q_t'\,O_t'(A^{-1})'.$$

A^{-1} is a unit-lower-triangular matrix. $\Lambda_t^{0.5}$ is a diagonal matrix of stochastic volatilities; the elements of the vector of logs of the diagonal elements of Λ_t, denoted $\log\lambda_t$, evolve as AR(1) processes with correlated errors:

$$\log\lambda_t = \Psi_0 + \Psi_1\log\lambda_{t-1} + e_t \quad \text{with} \quad e_t \sim N(0,\Phi),$$

where Ψ_0 is a vector of intercepts and Ψ_1 is a diagonal matrix of AR(1) coefficients.

The outliers enter the model in a diagonal matrix of scale factors, denoted O_t, with diagonal elements $o_{j,t}$ that are mutually $i.i.d.$ over all j and t. The outlier $o_{j,t}$ has a two-part distribution that distinguishes regular observations with $o_{j,t} = 1$ from outliers for which $o_{j,t} \geq 2$.[25] Outliers in variable j, $j = 1,\ldots,N$, occur with probability p_j and the

[25]The lower bound of 2 on the scale shift in outliers is motivated by seeing outliers as events firmly outside the typical mass of their otherwise Gaussian distribution (conditional on $o_{j,t}$).

distribution:

$$o_{j,t} = \begin{cases} 1 & \text{with probability} \quad 1 - p_j \\ U(2,20) & \text{with probability} \quad p_j, \end{cases}$$

where $U(2, 20)$ denotes a uniform distribution with support between 2 and 20. Stock and Watson (2016) first used this outlier specification with stochastic volatility in a multivariate unobserved component model of inflation.

Finally, the frequent-change volatility states consistent with fat-tailed innovations enter the model in the diagonal matrix Q_t with elements $q_{j,t}$ that are mutually $i.i.d.$ over all j and t. This component of the model is equivalent to having t-distributed VAR residuals (conditional on Λ_t and O_t). Following Jacquier, Polson, and Rossi (2004), we let the squares of the diagonal elements of Q_t, $q_{j,t}$, have inverse-gamma distributions:

$$q_{j,t}^2 \sim IG\left(\frac{d_j}{2}, \frac{d_j}{2}\right).$$

The j^{th} residual $q_{j,t} \cdot \varepsilon_{j,t}$ (adjusted for the rotation by A^{-1} and scaling by $\Lambda_t^{0.5} O_t$) has a student-t distribution with d_j degrees of freedom, since $\varepsilon_{j,t} \sim N(0, 1)$ and $d_j / q_{j,t}^2 \sim \chi_{d_j}^2$.

4.4 Model estimation

Our models are estimated with a Gibbs sampler, based on the methods of Carriero, Clark, and Marcellino (2019) (henceforth "CCM") for estimating large BVARs, but as corrected in Carriero, et al. (2022). As in CCM, we use a Minnesota prior for the VAR coefficients Π and follow their other choices for priors as far as applicable, too. Throughout, our BVARs include $p = 4$ lags.

For the infrequent outlier component of the BVAR-SVO-t model, we follow Stock and Watson (2016) in placing a beta prior on the outlier probability p_j. The prior is set to imply a mean outlier frequency of once every 10 years in quarterly data for SVO-t estimates, with precision set to be consistent with 10 years' worth of prior observations. For the

t-distributed component of the BVAR-SVO-t model, we follow Jacquier, Polson, and Rossi (2004) and estimate the degrees of freedom d_j for each variable using a uniform discrete prior with a range of 3 to 40. Carriero, et al. (2024) provide details on the Gibbs sampler steps for these and other volatility components of the model.

Parameter estimates from the BVAR-SVO-t model are based on 5,000 retained draws, obtained by sampling a total of 7,500 draws and discarding the first 2,500. For each retained draw of parameters in the Gibbs sampler, we produce 50 draws of forecasts.[26] Accordingly, all results in the paper are based on a total of 250,000 draws of forecasts from the BVAR's posterior predictive distribution.

5 Effect of SPF-tilting on model forecasts

This section evaluates the effects of entropic tilting when applied to predictive densities for real GDP growth, growth in non-residential fixed investment, growth in payroll employment, the unemployment rate, core PCE inflation, and the 5-year Treasury yield, comparing the predictive densities of the BVAR-SVO-t model against the available histogram forecasts reported by the SPF. As described in Section 4.1, as available we include SPF histograms for GDP growth, unemployment, and inflation, for horizons from the current year through three years ahead. While we do not tilt to survey forecast information for investment, employment, or the Treasury yield, tilting to forecasts of the other variables will impact the model's forecasts of these variables.

Entropic tilting is implemented using the analytical solution detailed in the preceding section. Of course, a single given histogram (one variable at one horizon) has multiple bins to hit. Histograms for additional variables or horizons adds to the bins to hit. The timing and horizons of BVAR forecasts are aligned with SPF survey rounds.

[26]This approach builds on the ideas of Waggoner and Zha (1999). At every Gibbs sampler node m, we employ J draws of the volatility components and VAR shocks over the forecast horizons, to better balance computational costs against a high degree of accuracy in the Monte Carlo approximation of the VAR's predictive likelihood.

Our tilting implementation appropriately bridges between the underlying quarterly forecasts from the BVAR and the annual forecasts of the SPF histograms. Specifically, starting from the quarterly draws, we perform the necessary transformations to obtain BVAR draws of calendar year forecasts as defined by the SPF, to which we apply the tilting targets.[27]

5.1 Visual fit to SPF histograms

We begin by illustrating the differences between raw model output and SPF histogram forecasts for selected forecast origins, of 2009Q1 and 2021Q2. In both instances, reflecting somewhat unusual economic circumstances associated with the Great Recession (2009Q1) and the recovery following the COVID-19 pandemic (2021Q2), the SPF forecasts differ from the BVAR forecasts, such that entropic tilting to the SPF histograms has an impact on the model's forecasts.

Figures 5 through 8 provide predictive densities for annual forecasts corresponding to SPF horizons, from the current year through two years ahead (we omit three years ahead for chart readability). Separate figures provide results for targeted variables (those for which SPF histograms are available) and non-targeted variables (for which histograms are not available). Each panel provides the BVAR forecasts as red (untilted) and black (tilted) lines. For variables and horizons for which SPF histograms are available, the relevant panels show the SPF histograms with black bars and untilted BVAR histograms with red bars.

Starting with basic features of the SPF histograms and the BVAR's untilted forecasts, the baseline model, the BVAR-SVO-t specification, produces smooth predictive densities. In contrast, the SPF histograms in some instances exhibit mass concentrated in specific bins, reflecting survey participants' discrete expectations. Most starkly, in 2009Q1, when

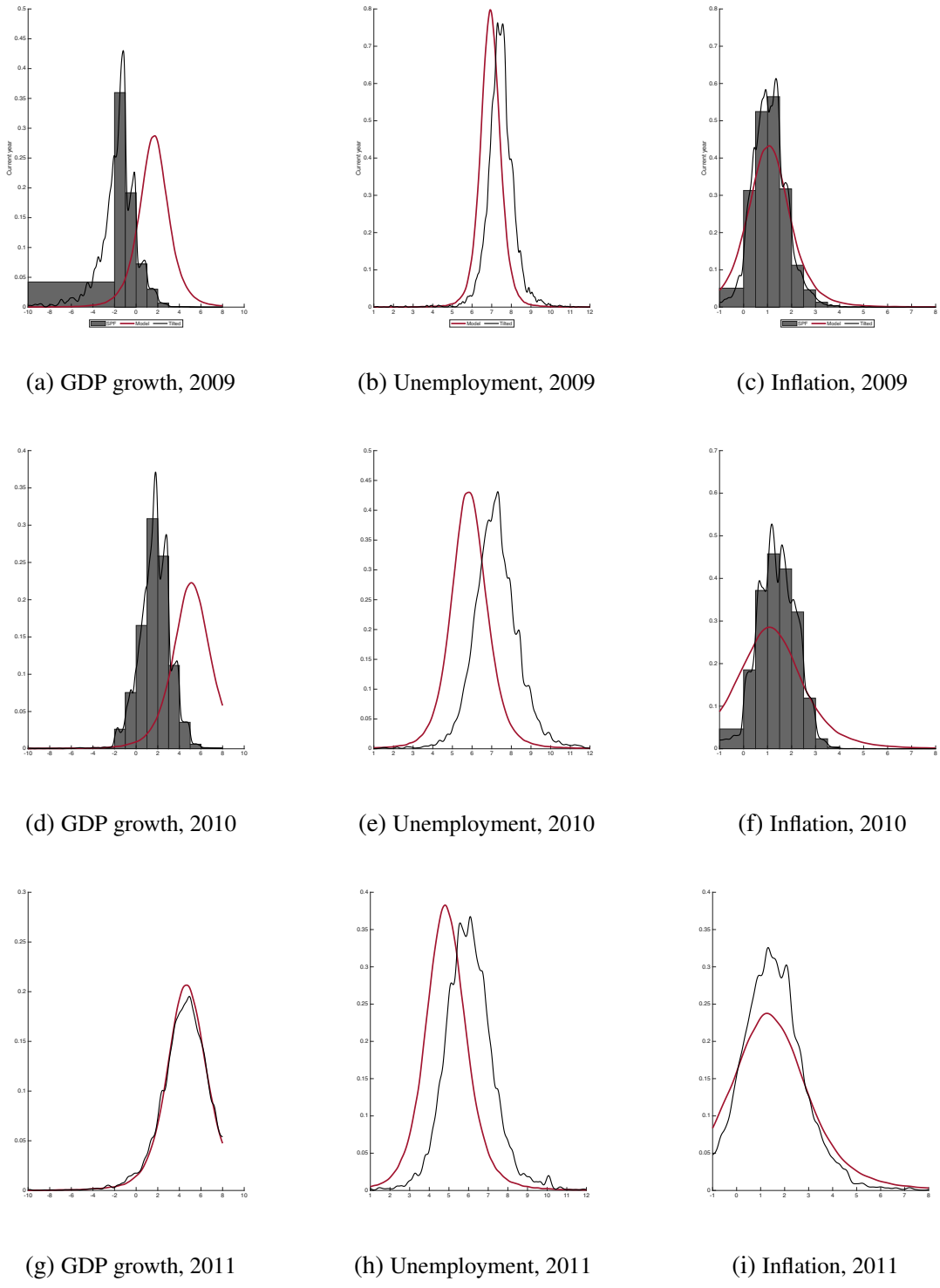
[27] Reflecting the SPF's variable-specific convention, these transformation can be linear (e.g., taking annual averages of quarterly forecasts of the unemployment rate) or nonlinear (e.g. calculating growth rates of annual average levels of GDP).

the economy was still in recession, the SPF predictive density for GDP growth in 2009 was sharply skewed to the left, with substantial probability on GDP growth of -2 percent or less and an 89 percent probability of GDP declining in 2009 (panel 5a). While the BVAR's predictive density shows a significant chance of a decline in GDP for the year, the model had a much more favorable outlook for GDP growth. In this instance, tilting the BVAR's forecasts to the SPF histograms substantially shifts the model's predictive density for GDP growth in 2009 to the left and sharpens the density's mode. In general, though, as noted above, tilting the model output to match the SPF histogram redistributes probability mass across bins, preserving the overall shape while aligning with survey responses.

Tilting to the SPF histograms also impacts the model's forecasts for unemployment and core PCE inflation (in 2009Q1, there were not yet published unemployment histograms to incorporate as tilting targets), as well as the forecasts for employment, investment, and the 5-year Treasury yield. As expected, with tilting reducing the GDP growth forecast, it also raises the unemployment rate forecast, shifting the model's predictive density to the right (panel 5b), and significantly reduces the forecasts of growth in employment (panel 6a) and investment (panel 6b). The available SPF forecast of core inflation tends to reduce the dispersion of the BVAR's forecast for inflation in 2009 (panel 5c). The tilting of the BVAR's forecasts to SPF histograms for GDP growth and inflation also leads to some leftward shift in the forecast of the 5-year Treasury yield (panel 6c).

Forecasts for the year-ahead horizon, the year of 2010, show patterns similar to those for the current year forecasts, although with tilting in this instance leading to visually sharper shifts in the BVAR's predictive densities and also tending to shift to the right the model's forecast for core PCE inflation (panel 5f). The BVAR's forecasts for two years ahead continue to show some impact of tilting to SPF histograms, but smaller impacts than for shorter horizons (panels 5g, 5h, 5i, 6g, 6h, and 6i).

Figure 5: Targeted variables, tilted and untilted predictive densities in 2009Q1



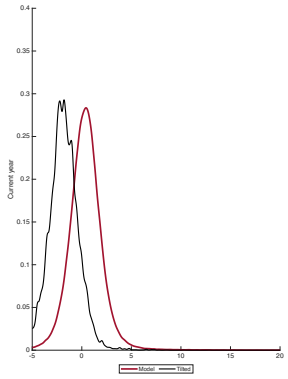
(a) GDP growth, 2009 (b) Unemployment, 2009 (c) Inflation, 2009

(d) GDP growth, 2010 (e) Unemployment, 2010 (f) Inflation, 2010

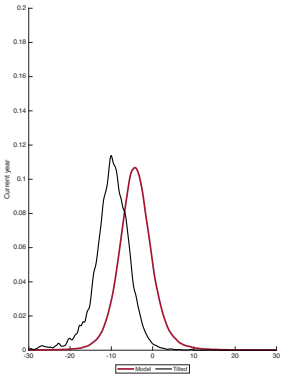
(g) GDP growth, 2011 (h) Unemployment, 2011 (i) Inflation, 2011

Note: Model-based densities in red, tilted densities and SPF histograms in black.

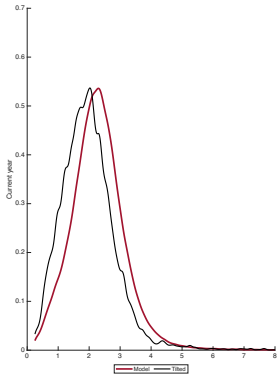
Figure 6: Non-targeted variables, tilted and untilted predictive densities in 2009Q1



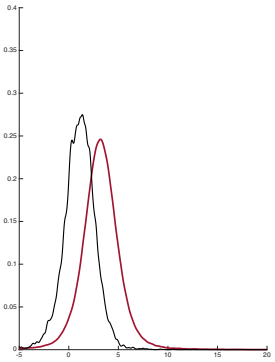
(a) Payroll employment, 2009



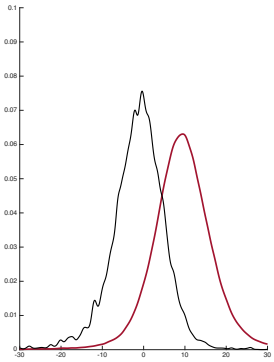
(b) Investment, 2009



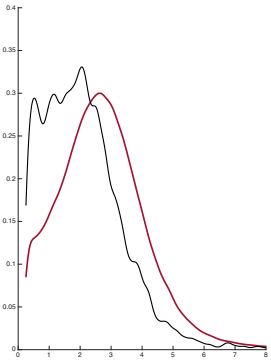
(c) 5-year yield, 2009



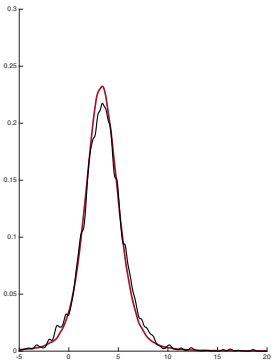
(d) Payroll employment, 2010



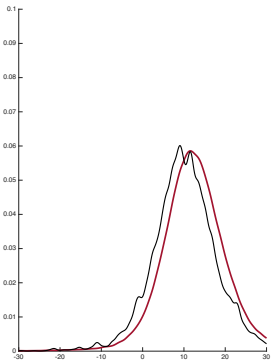
(e) Investment, 2010



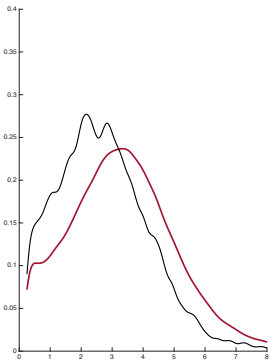
(f) 5-year yield, 2010



(g) Payroll employment, 2011



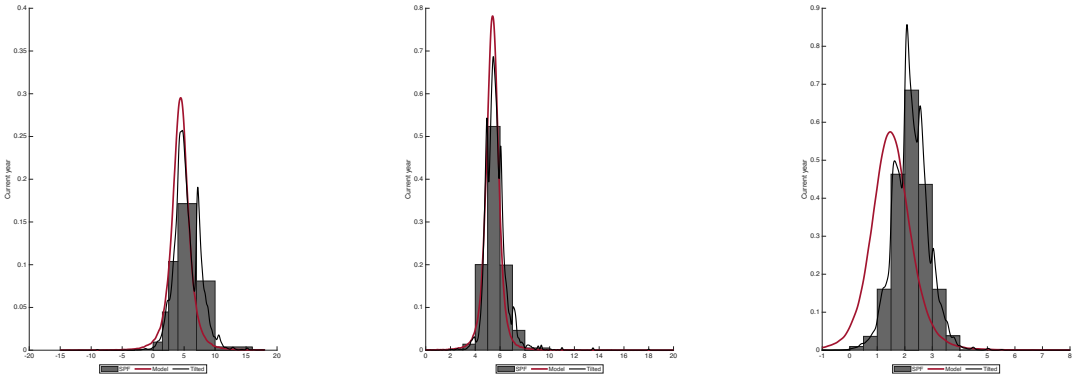
(h) Investment, 2011



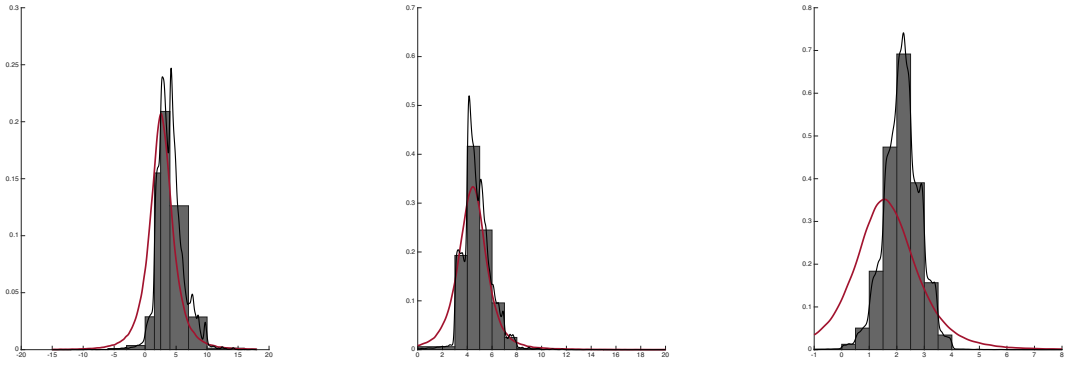
(i) 5-year yield, 2011

Note: Model-based densities in red, tilted densities and SPF histograms in black.

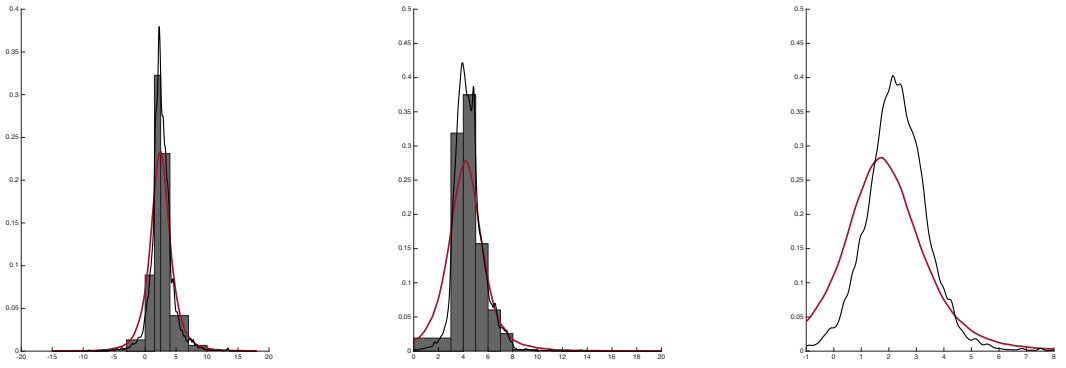
Figure 7: Targeted variables, tilted and untilted predictive densities in 2021Q2



(a) GDP growth, 2021 (b) Unemployment, 2021 (c) Inflation, 2021



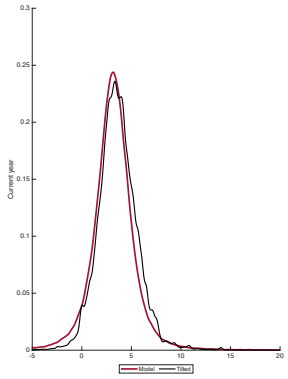
(d) GDP growth, 2022 (e) Unemployment, 2022 (f) Inflation, 2022



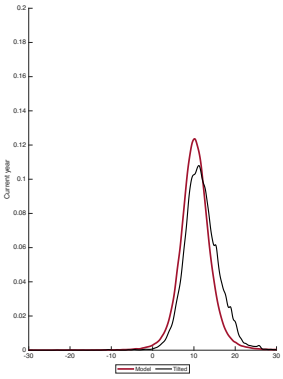
(g) GDP growth, 2023 (h) Unemployment, 2023 (i) Inflation, 2023

Note: Model-based densities in red, tilted densities and SPF histograms in black.

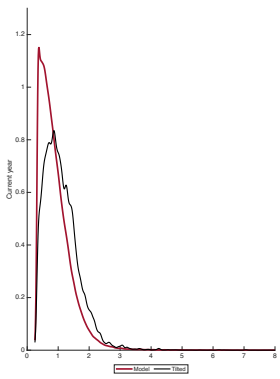
Figure 8: Non-targeted variables, tilted and untilted predictive densities in 2021Q2



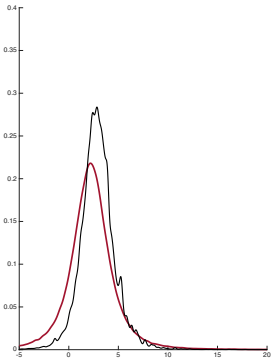
(a) Payroll employment, 2021



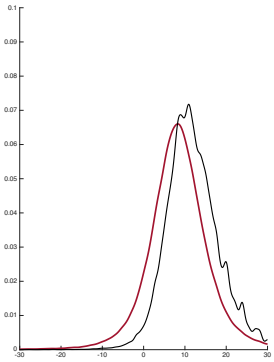
(b) Investment, 2021



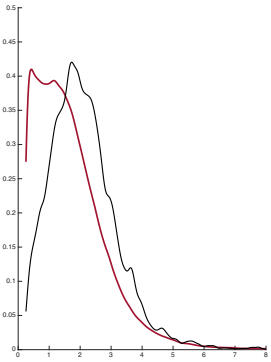
(c) 5-year yield, 2021



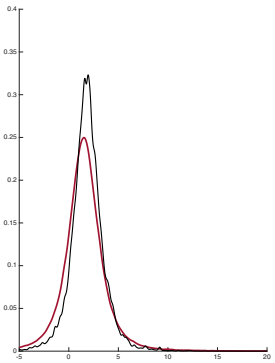
(d) Payroll employment, 2022



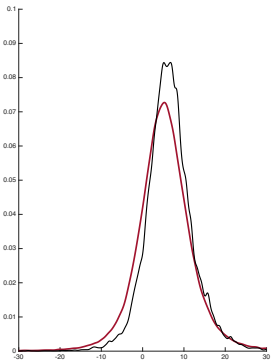
(e) Investment, 2022



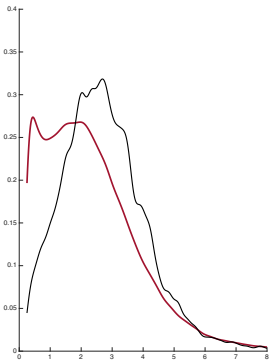
(f) 5-year yield, 2022



(g) Payroll employment, 2023



(h) Investment, 2023



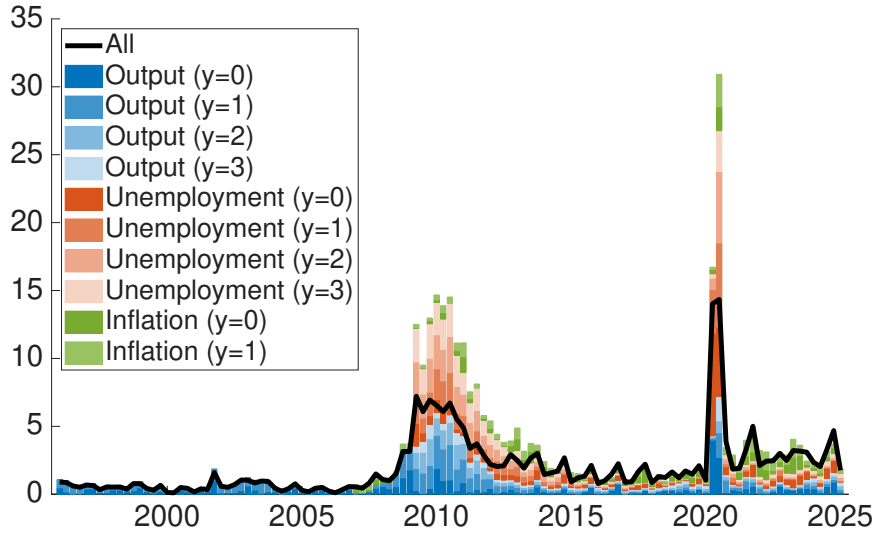
(i) 5-year yield, 2023

Note: Model-based densities in red, tilted densities and SPF histograms in black.

In 2021Q2, with the economy continuing to recover from the COVID-19 pandemic, the SPF predictive density for GDP growth in 2021 is somewhat skewed to the right, as shown in Figure 7. The histogram's mode put a probability of 51 percent on growth between 4 and 6.9 percent and a probability of 24 percent on growth between 7 and 9.9 percent, compared to a probability of 16 percent on growth between 2.5 and 3.9 percent. The SPF's density forecasts for the unemployment rate and core PCE inflation in 2021 appear largely symmetric, but fairly dispersed. The SPF histogram forecasts for 2022 show somewhat more skew, most noticeably to the right for both GDP growth and unemployment. Entropically tilting the BVAR's forecasts to match these histograms results in predictive densities for GDP growth with two peaks evident for growth in both 2021 and 2022 (panels 7a and 7d) and inflation in 2021 (panel 7c). Such multiple peaks can arise inherently from tilting to match histogram targets while leaving mass distributed within bins as in the BVAR's density so as minimize the distance between the tilted and untilted BVAR densities. Tilting to the SPF forecasts shifts the BVAR's predictive distributions of core PCE inflation in 2021 and 2022 noticeably higher (panels 7c and 7f). The incorporation of information in the SPF histograms also shifts up the predictive densities for employment, investment, and the 5-year Treasury yield, somewhat more evidently for 2022 (panels 8d, 8e, and 8f) than 2021 (panels 8a, 8b, and 8c). At the two year-ahead horizon, the impacts of tilting to the SPF histograms are more modest than at shorter horizons, except that clear rightward shifts of the predictive density persist for core inflation and the 5-year Treasury yield (panels 7i and 8i).

As the preceding examples indicate, at times entropic tilting of the BVAR's forecasts to SPF histograms can materially alter the model's forecasts in order to closely follow the SPF bin probabilities. To provide a broader time perspective on how much such tilting impacts the baseline model forecasts, we examine the time series of the KLIC distance between the untilted and tilted distributions from 1996 into 2025. The KLIC distance provides a measure of how much (at each forecast origin) the conditions provided by

Figure 9: KLIC decomposition over time



Note: KLICs from entropic tilting at different forecast origins. 'All' refers to our baseline set of targets, and the colored bars indicate the KLICs obtained from targeting a given individual histogram.

the SPF histograms alter the baseline BVAR's predictive density. The model's ability to adapt to these conditions is evident in the tilting process, which effectively redistributes probability mass to align with the survey data.

Figure 9 provides the time series of total KLIC distance (black line) and contributions (colored bars) from tilting to each available histogram of the set of targets. The contributions are defined by the first term of equation (32)'s decomposition of relative entropy, which sums the relative entropies between the marginal distributions for each histogram.[28] As shown in the figure, the information content conveyed by the SPF (as measured by the KLIC) has varied over time. In the period of (relatively speaking, by historical standards) macroeconomic stability from 1996 through 2006 — a period in which the only SPF histograms available for tilting were current- and next-year GDP growth — the KLIC distance between the model's raw forecast and the tilted forecast density was relatively small and stable, with varying (over time) contributions from both the current-

[28]The sum of these contributions differs from the reported total KLIC due to unreported contributions associated with adjustments in tilting weights necessary to preserve the proposal's correlation structure as much as possible.

and next-year GDP forecasts. Subsequently, the information content conveyed by the SPF (as measured by the KLIC) rose particularly around the GFC and its aftermath for a few years and then again during the COVID-19 pandemic. Around the GFC, the relatively high level of entropic distance reflects sizable contributions from the tilting to the SPF histograms for both GDP growth and the unemployment rate (across multiple horizons), with a much smaller contribution from inflation. With COVID's outbreak, the growth and unemployment histograms again make sizable contributions to the KLIC distance in predictive distributions. The inflation histograms have a larger impact on the KLIC distance and tilted distribution in the few years after the pandemic's outbreak, a period that featured elevated inflation.

5.2 Forecast accuracy evaluation

We assess the performance of the tilted and untilted forecasts using real-time forecast evaluations, of forecasts made starting in 1996Q1 and continuing into 2025. As indicated above, drawing on the histogram forecasts available in the SPF, our tilting targets include GDP growth, unemployment, and core PCE inflation. For each variable, we use all of the SPF horizons available, which vary over time and by variable. In the case of GDP growth, we use the current year and next-year forecast histograms in the entropic tilting starting with 1996Q1 and (in addition) the forecast histograms for two and three years ahead starting with the 2009Q2 forecast origin. For the unemployment rate, we use all of the SPF histograms for horizons from the current year through three years ahead, starting with the 2009Q2 forecast origin. For core PCE inflation, we use the current and next year forecast, starting with 2007Q1. Of course, one could consider tilting the BVAR forecasts to other, more limited combinations of the available SPF histogram forecasts, and we will include such comparisons in our empirical analysis. We view our chosen baseline including a larger set of targets to have the advantage of being comprehensive in capturing the SPF's view of the macroeconomic outlook and providing a simple and broad look at

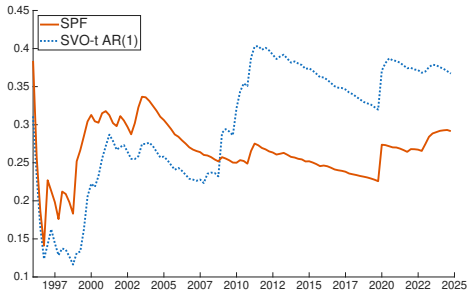
the potential efficacy of information BVAR forecasts with survey-based density forecasts in the form of the SPF histograms.

5.3 Forecast accuracy of annual histogram predictions

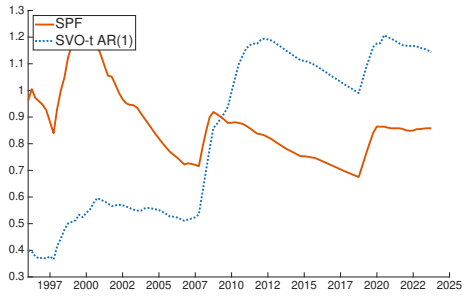
We begin by gauging the potential for improved forecast accuracy from tilting to the SPF by comparing the SPF histogram forecasts to the BVAR's predictive densities for corresponding annual forecasts. That means we compare the quality of predictions for outcomes to fall into a given SPF-defined bin as predicted by the model (without tilting) and the SPF histogram forecasts. As summary measure, we computed discrete rank probability scores (DRPS) for both the BVAR-SVO-t model and the SPF histogram forecasts. The DRPS is defined as $\text{DRPS}_t = \sum_{k=1}^{K} \left(P_t^k - D_t^k\right)^2$, with K indicating the number of bins, P_t^k denoting the cumulative forecast probability from bins 1 through k, and D_t^k is an indicator that equals 1 if the outcome falls in any bin equal to or lower than k and 0 otherwise. The lower the score, the better the forecast. Past studies of survey forecasts using the DRPS include Boero, Smith, and Wallis (2011), Clements (2018), and Krueger and Pavlova (2024).

The real-time contributions of realized outcomes to the DRPS score are shown in Figure 10, for the samples over which annual forecasts from the SPF are available at a given horizon. For chart readability, we report results for the current year and year-ahead horizons; results for longer horizons are qualitatively similar. As can be seen from the figure, for outcomes prior to the Great Recession (i.e. forecast origins prior to 2006), the SPF's density forecasts of GDP growth at the current- and next-year horizons were less accurate than the BVAR's annual forecasts (on average, by roughly 20 to 40 percent). However, for the period since the Great Recession, the SPF's density forecasts are generally more accurate than the BVAR's. For real GDP growth, the SPF's advantage is sizable (roughly 30 to 40 percent) and consistent over the period since the Great Recession. In the case of the unemployment rate and core PCE inflation, the SPF's forecasts show the

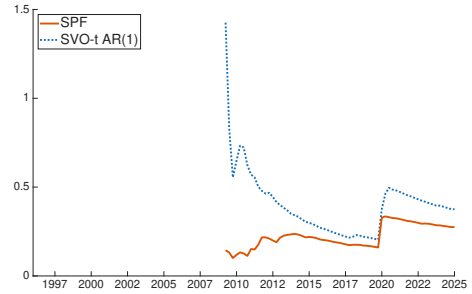
Figure 10: DRPS for forecasts from BVAR-SVO-t model and SPF



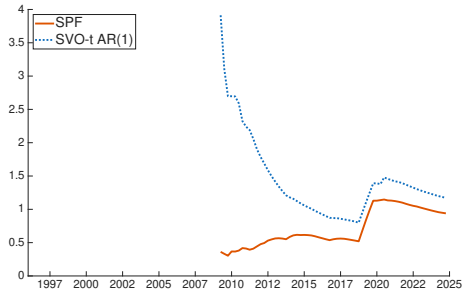
(a) GDP growth, current year



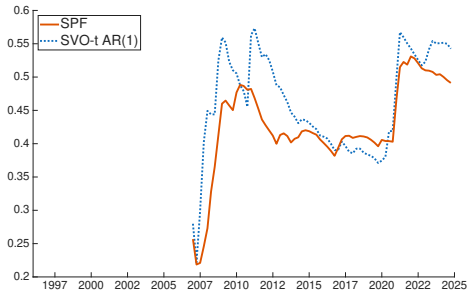
(b) GDP growth, next year



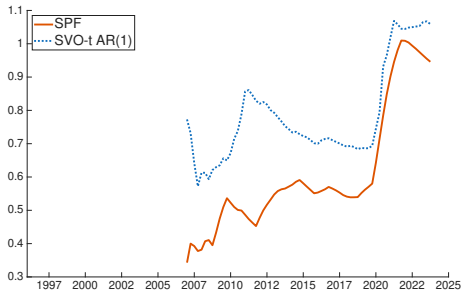
(c) Unemployment rate, current year



(d) Unemployment rate, next year



(e) Core PCE inflation, current year



(f) Core PCE inflation, next year

Note: The figure shows time-averaged discrete rank probability score (DRPS) computed over growing windows. DRPS are shown for the BVAR-SVO-t model and SPF histogram forecasts. The DRPS is calculated for current year and one-year-ahead forecasts, respectively. Lower values indicate better density-forecast performance. The x-axis represents the forecast horizon, while the y-axis indicates the cumulative average of the relative probability score.

largest average gains relative to the BVAR through about 2016 and then smaller gains for the remainder of the sample (with average gains over the entire sample of about 20 percent for unemployment and 10 percent for inflation). Overall, these DRPS results for annual density forecasts from the SPF compared to the BVAR indicate that, since the Great Recession, the SPF histogram forecasts provide valuable information for forecasting GDP growth, unemployment, and inflation.

5.4 Forecast accuracy of BVARs tilted to SPF histograms

From this evidence of predictive content in the SPF's histograms, we turn to examining the potential forecast accuracy gains that may be achieved by using entropic tilting to inform BVAR forecasts with SPF histogram forecasts. More specifically, we compare tilted BVAR forecasts to an untilted baseline in terms of forecasts of GDP growth, investment growth, employment growth, unemployment, inflation, and the 5-year yield at horizons ranging from horizons of $h = 0$ through 15 (so from the current quarter through four years ahead). In these results, the unemployment rate and Treasury yield forecasts are quarterly levels, and other forecasts are expressed as four-quarter growth rates. Table 1 provides results for the full sample of 1996-2025. For each variable and horizon, the table reports the ratio of the RMSE or CRPS for the tilted forecast to the corresponding score for the untilted BVAR-SVO-t forecast, with values below one indicating improved performance from tilting. The tables also indicate statistical significance of the relative RMSE and CRPS as measured by Diebold-Mariano-West tests.

These results indicate that, on average over the full sample, entropically tilting BVAR forecasts to the annual SPF histograms consistently improves the BVAR's quarterly forecast accuracy, both point (RMSE) and density (CRPS). Accuracy is improved for not only those variables for which histogram targets are available (GDP growth, unemployment, and core inflation) but also those variables without targets (growth in employment and investment and the 5-year Treasury yield). Otherwise, the gains vary by variable and horizon, but are

generally smallest for the Treasury yield.[29] The gains yielded by tilting to SPF histograms can be especially large at the horizon $h = 0$ — e.g., improving the RMSE accuracy of GDP and unemployment forecasts by roughly 50 to 70 percent and the accuracy of investment and employment forecasts by about 20 to 40 percent. At horizons between $h = 4$ and $h = 8$, the tilted forecast's accuracy gains are roughly on the order of 15 to 30 percent for GDP, unemployment, employment, and inflation; 10 percent for investment; and 5 percent for the Treasury yield. Accuracy improvements persist at longer horizons, although by smaller magnitudes for some variables (e.g., GDP and unemployment). While omitted in the interest of brevity, corresponding results for a sample ending before the pandemic are qualitatively very similar.

While the full sample shows gains (on average) to our proposed tilting approach, the benefits of incorporating the information from SPF histograms could vary over time. One specific driver of time variation could be the changing availability of SPF histograms for tilting. As noted above, before 2007 the tilting impacts are driven entirely by the information in the current- and next-year GDP forecast histograms. SPF histograms for inflation and unemployment only impact the tilted forecasts starting in 2007 and 2009, respectively. Of course, in addition, the DRPS results above showed SPF histogram forecasts of annual GDP growth before 2007 to be less accurate than model-based forecasts. Accordingly, Table 2 provides quarterly forecast accuracy results for a sample of 2007-2025. Broadly, the RMSE and CRPS ratios are qualitatively the same for the 2007-2025 and 1996-2025 samples. Consistent with the previous evidence on DRPS accuracy before 2007, compared to the full sample results the shorter sample yields more instances of statistical significance and for some variables and horizons slightly larger gains in accuracy. It follows that the benefits of entropic tilting BVAR-based forecasts to SPF forecast histograms are driven by the period since 2007, corresponding to the period since the Great Recession. With only GDP growth histogram forecasts available for tilting from 1996

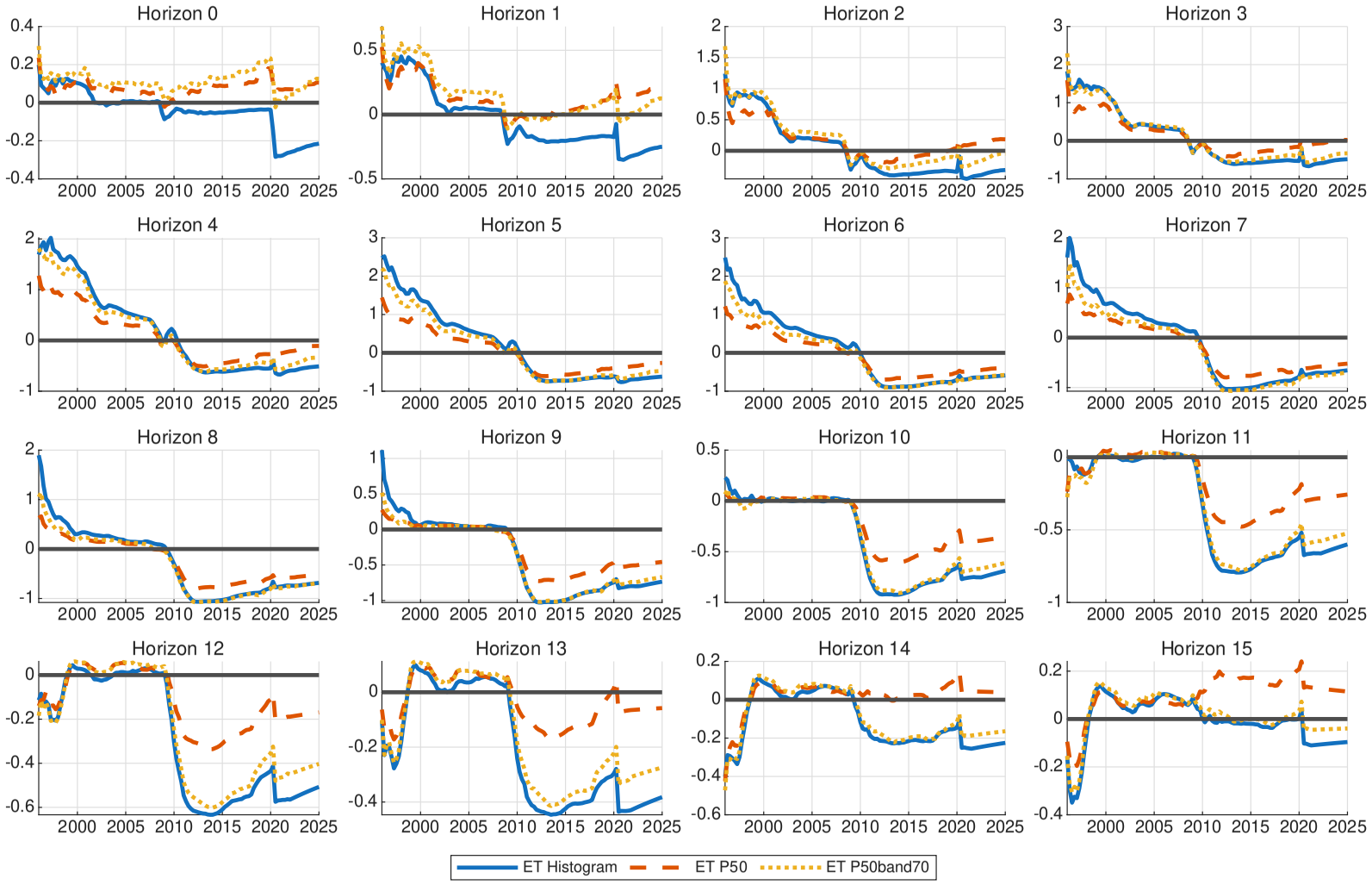
[29]The finding of smaller findings for the Treasury yield may be seen as broadly consistent with the term structure literature in which, for interest rate forecasting, a random walk is hard to beat.

through 2006, there is little implied benefit to bringing the SPF densities to bear in this pre-Great Recession period.

The benefits of tilting to SPF histograms could also vary within the period since the Great Recession. For example, the judgment of forecasters captured in the SPF might be especially helpful around business cycle turning points and less helpful in the middle of economic expansions. As shown above, the KLIC distance between the untilted and tilted predictive densities was greatest around the GFC and the COVID-19 pandemic and smaller in other periods. To examine time variation in relative accuracy, we consider scores of the untilted and tilted forecasts computed over an expanding time sample as forecasting moves forward and report time series of the difference in the cumulative average scores. To compact results and provide a measure of overall density accuracy, the score measure is specified as the energy score (ES), a multivariate generalization of the CRPS. The blue lines in Figure 11 provide the time series of the difference in the cumulative average scores of the forecast tilted to all available histograms and the BVAR's forecast, with negative entries indicating the tilted forecast to be more accurate.

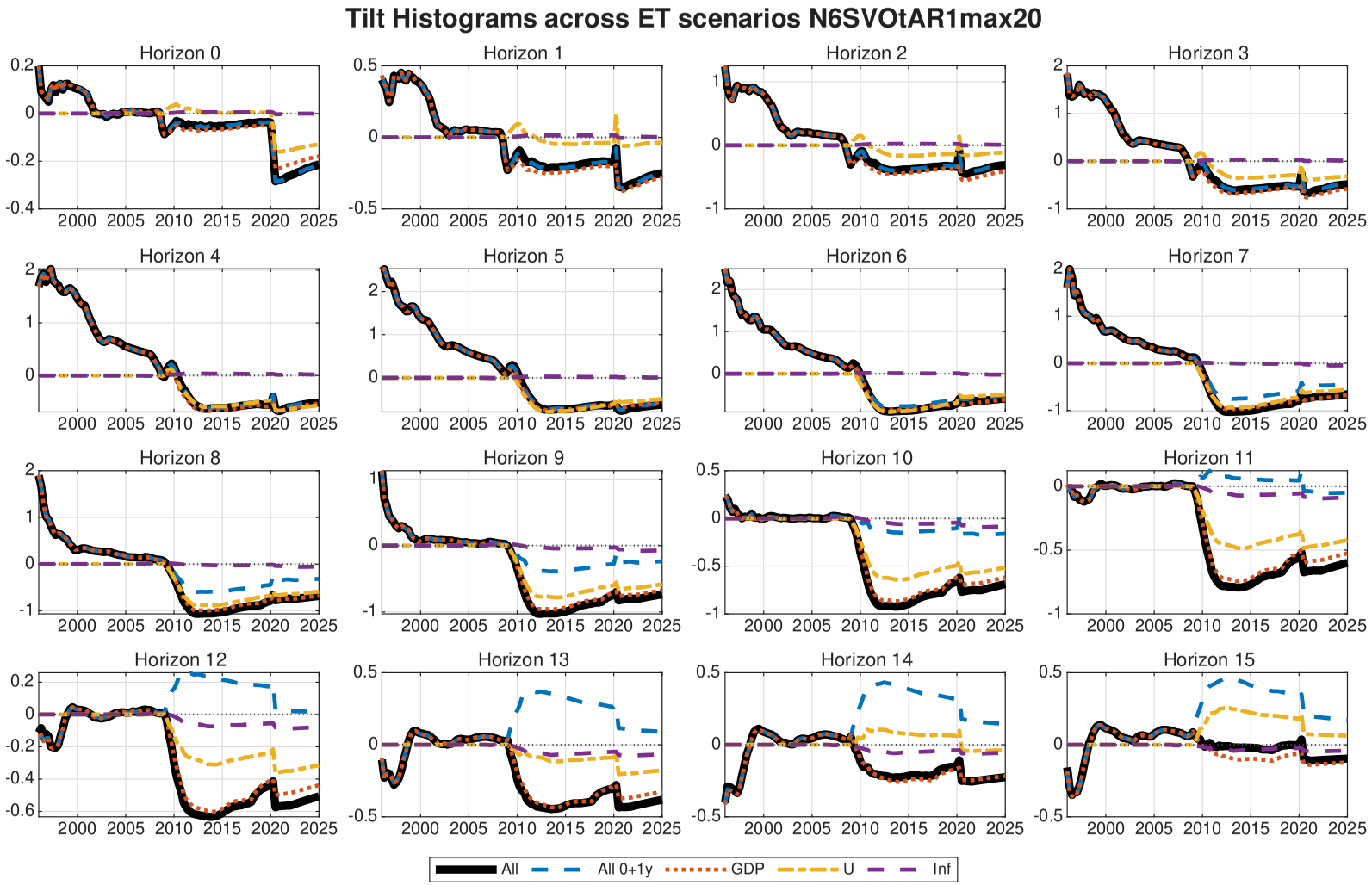
Consistent with the comparison of the average results in the tables for the 1996-2025 and 2007-2025 samples, the time series of cumulative average ES differences show little consistent advantage to tilting to just the SPF's current- and next-year GDP growth histograms in roughly the first 10 years of the sample. Rather, for most horizons, the BVAR forecast is more accurate than the tilted forecast. Following the onset of the Great Recession in late 2007, the cumulative average loss differences dropped, often sharply and were then sustained for most of the remainder of the sample. This indicates the judgment of professional forecasters embedded in the SPF histograms was especially helpful following the Great Recession. Plots of time series of the BVAR's median forecast with the SPF histogram-implied mean (omitted in the interest of brevity) show that, in this period, the SPF forecasts improve accuracy by moderating the model's too-strong forecasts of GDP growth and the unemployment rate, while tending to edge up the model's

Figure 11: Accuracy of tilted to untilted density forecasts over time



Note: Differences between cumulative average energy scores between tilted and untilted densities generated from BVAR-SVO-t. Negative values indicate better scores of tilted densities.

Figure 12: Forecast accuracy when tilting to different targets over time



Note: Differences between cumulative average energyscores between tilted and untilted densities generated from BVAR-SVO-t and using alternative tilting targets. Negative values indicate better scores of tilted densities.

relatively low forecasts of core inflation. With the pandemic, the relative score results show some additional gains to tilting to SPF.

5.5 Tilting to a few histogram-consistent percentiles

With previous studies such as those mentioned in the introduction finding that point forecasts from the SPF can be used to improve forecasts from time series models, a natural question is whether the information content we find in SPF histograms is driven mostly or entirely by matching the center of the histograms, or whether there is information in the entire histogram. To assess this question, we apply our analytical solution to entropic tilting to tilt the BVAR's forecasts to just the medians of the SPF's predictive histograms. In this comparison, the set of variables and forecast horizons in the tilting is the same as in our baseline tilting to complete SPF histograms. The histogram-implied percentiles are calculated as the midpoints of the histogram bins where the cumulative probability reaches or exceeds the desired percentile level. For example, if the cumulative probability reaches 0.50 within a bin, the median is taken as the midpoint of that bin. If the cumulative probability jumps from below 0.50 in one bin to above 0.50 in the next bin, the median is taken as the midpoint of the second bin.

As a joint measure of forecast accuracy, the red dotted lines in Figure 11 provide time series of the difference in the cumulative average scores of the forecast tilted to all available histogram medians and the BVAR's forecast (again, negative entries indicating the tilted forecast to be more accurate). These results indicate tilting to just medians to be less effective than tilting to full histograms. At shorter horizons, the medians-tilted forecast is less accurate than the BVAR for nearly all of the sample. At longer horizons, the medians-tilted forecast typically improves (since 2007) on the BVAR's accuracy, but by less than the baseline approach of tilting to full histograms.

The upper panel of Table 3 provides RMSE and CRPS accuracy comparisons for each variable over the 1996-2025 sample (results for 2007-2025 are qualitatively the same). The

panel compares the forecasts with tilting to implied 50th percentiles of SPF histograms with the untilted BVAR forecasts. (To save space, we omit a few longer forecast horizons.) Broadly, the gains to entropic tilting to 50th percentiles reported in this table panel are smaller and more selective than the gains shown in Table 1's baseline results. Tilting to the SPF histogram's medians has some benefit (relative to the baseline BVAR) for unemployment and employment at short horizons and for inflation at most horizons, while harming the RMSE accuracy of GDP and investment growth forecasts at short horizons and the RMSE accuracy of Treasury yield forecasts at most horizons. In contrast, as noted earlier, tilting to the full histograms yields consistent and more sizable gains.

A related question is whether most of the information in the SPF histograms could be captured by the center of the histograms coupled with some limited information on their shape. For this case, we tilt the BVAR forecasts to three percentiles of full set of available SPF histograms: 15th, 50th, and 85th. This set of moments captures some rough information about the centers and shapes of the histograms (widths and possible asymmetries).

The yellow dotted lines in Figure 11 show relative ES performance for this tilting approach as compared to the baseline of tilting to full histograms and the more limited alternative of tilting to histogram-consistent medians. By this joint accuracy measure, the relative accuracy of tilting to the median and 70 percent band falls in between that achieved by tilting to all histograms and tilting to just medians. This in-between approach can yield gains in line with those achieved by using complete histograms but does not do so consistently. At shorter horizons, forecasts tilted to the 15th, 50th, and 85th percentiles are less accurate than the BVAR for nearly all of the sample. But at many longer horizons (though not all) this tilting strategy yields gains comparable to those achieved with tilting to full histograms.

The reported joint density accuracy translates into comparable results for RMSE and CRPS accuracy by variable, provided (for the 1996-2025 sample) in the lower panel of

Table 3. At most horizons shown, tilting the BVAR's forecasts to this set of histogram percentiles improves on tilting to just the histogram median. On average over the full sample, for most horizons this tilting to 15th, 50th, and 85th percentiles yields consistent, modest to moderate gains in accuracy for forecasts of each variable, with exceptions for some h = 0 and h = 15 forecasts. Based on these results, the shape of SPF histograms appears to have some information content for BVAR forecasts. However, as conveniently summarized by the energy scores, tilting to the full histograms yields more consistent (across horizons and time in some cases) gains than does tilting to the 15th, 50th, and 85th percentiles of the histograms.

5.6 Tilting to selected SPF histograms

As noted above, while we have focused on tilting BVAR forecasts to a relatively comprehensive set of SPF histograms, one could instead choose to tilt to subsets of horizons or variables on the basis of subjective preference or perhaps performance. Many possible permutations are possible. Rather than engage in an exhaustive search and set of results, we consider a basic range of alternative sets of tilting targets, in which the BVAR forecasts are tilted to: (i) all of the available SPF forecasts for growth, unemployment, and inflation for the current and next year; (ii) all of the available SPF forecasts of GDP growth (covering all horizons); (iii) the same (i.e., all forecast horizons available) for unemployment, and (iv) the same for core PCE inflation. To be clear, when we tilt to forecasts of single variable, we tilt all of the BVAR's forecasts (i.e., all variables) to those single-variable histograms.

Starting with the ES measure of joint density accuracy, Figure 12 provides time series of the difference in the cumulative average scores for tilted versus untilted forecasts under the baseline set of histogram targets and the four alternative sets. The approach of tilting to just current and next year histograms (blue dotted lines) generally rivals the baseline approach's accuracy at shorter horizons (from h = 0 to 6) but falls short at longer horizons; at horizons of 11 or more quarters, the BVAR's forecasts are more accurate than those tilted

to current and next year histograms. In ES accuracy, tilting to all available GDP histograms (red dotted lines) seems to match the success of tilting to all histograms, except not quite matching in the post-pandemic period. Tilting to just unemployment histograms (yellow dotted lines) can improve on the BVAR's forecast accuracy, but typically by less than the baseline and GDP-only alternatives. The relative performance of tilting to unemployment or current and next year histogram alternatives varies with the forecast horizon. Finally, in joint density accuracy, tilting to just inflation histograms (purple dotted lines) is much less successful than the baseline and other tilting choices, yielding only small gains in accuracy relative to the BVAR at longer horizons and no gains at shorter horizons.

To assess efficacy for individual variables, Table 4 provides RMSE and CRPS results for the alternative choices of tilting targets comparisons, for the full sample of 1996-2025 (results for 2007-2025 are qualitatively similar and available in the online appendix). In this table, all tilting forecasts are compared against the raw BVAR forecasts, and to facilitate comparisons, results for the baseline tilting set of all variables is provided in the first column of results. Again, to save space, results are reported for a selected subset of horizons. Consistent with the ES results, these RMSE and CRPS results indicate that some alternative choices of tilting targets can achieve accuracy gains in line with those with the baseline of using all available histograms, but no other choice matches the baseline for its consistency across variables and horizons and magnitudes of gains. Tilting to just current and next year forecasts is close to the all-histograms baseline for most horizons and variables but not quite as accurate for some variables at longer horizons. Tilting to the histograms for a single variable typically matches the accuracy of the baseline tilting set for that variable but does not improve as much the accuracy of other variables. For one example, tilting to just GDP histograms improves unemployment rate forecasts but by less than the baseline tilting. For another, tilting to inflation histograms does not really improve forecasts of the unemployment rate or employment, unlike the baseline tilting specification. All together, these results indicate that while tilting to a more restricted set of histograms

can selectively improve on the accuracy of BVAR forecasts, there is no uniform advantage in forecast accuracy over the baseline choice of tilting to all available histograms.

6 Conclusion

This paper develops a new, direct approach to combining information in the density forecasts of the SPF with model-based forecasts. Taking the SPF histograms as given — without needing to fit histogram distributions or first and higher-order moments — our approach directly tilts the predictive distribution of time series model-based forecasts to the SPF histogram bins. Our solution is applicable to simulated densities from variety of time series models, including linear VARs, Quantile VARs and DSGE models

We first develop a new analytical characterization for the tilting problem when targeting histogram bins. To examine the empirical merits of the approach, we then examine the information content of SPF histogram forecasts for the accuracy of forecasts of major macroeconomic aggregates for the U.S. produced by a Bayesian VAR with time-varying volatility.

Using our analytical solution for the tilting problem, we entropically tilt the BVAR's quarterly forecast distributions to SPF histograms, including multi-year forecast horizons from the SPF for GDP growth, the unemployment rate, and core PCE inflation. Relative to the baseline model forecasts, SPF histogram forecasts add the most information content around the Great Recession and its aftermath, as well as during the COVID-19 pandemic, with most of the information coming from the SPF's GDP growth and unemployment rate forecasts. On average over samples of 1996-2025 and 2007-2025, entropically tilting BVAR forecasts to the annual SPF histograms consistently improves the BVAR's quarterly forecast accuracy (point and density). The benefits of entropically tilting the BVAR forecasts to the SPF histograms are sharpest in the years around the Great Recession and around COVID-19's outbreak in 2020 but persist for the rest of the post-2007 period.

Our framework for tilting model-based forecasts to histogram forecasts from an external can readily be applied with other sources of probabilistic forecasts. For example, in future applications we intend to extend the set of tilting targets to incorporate the SPF's recession-probability forecasts.

References

Adrian, Tobias, Domenico Giannone, Matteo Luciani, and Mike West (2025), "Scenario synthesis and macroeconomic risk," Manuscript, March 28.

Allayioti, Anastasia, Rodolfo Arioli, Colm Bates, Vasco Botelho, Bruno Fagandini, Luís Fonseca, Peter Healy, Aidan Meyler, Ryan Minasian, and Octavia Zahrt (2024), "A look back at 25 years of the ECB SPF," ECB Occasional Paper Series 364, European Central Bank.

Antolín-Díaz, Juan, Ivan Petrella, and Juan F. Rubio-Ramírez (2021), "Structural scenario analysis with SVARs," *Journal of Monetary Economics*, 117, 798–815.

Banbura, Marta, Federica Brenna, Joan Paredes, and Francesco Ravazzolo (2024), "The why and the how of combining VAR and survey density forecasts," working paper, SSRN.

Bassetti, Federico, Roberto Casarin, and Marco Del Negro (2022), "A Bayesian approach to inference on probabilistic surveys," Staff Reports 1025, Federal Reserve Bank of New York.

——— (2023), "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 Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw: Academic Press, chap. 15, 443–476.

Bobeica, Elena, and Benny Hartwig (2023), "The COVID-19 shock and challenges for inflation modelling," *International Journal of Forecasting*, 39, 519–539.

Boero, Gianna, Jeremy Smith, and Kenneth F. Wallis (2011), "Scoring rules and survey density forecasts," *International Journal of Forecasting*, 27, 379–393.

Carriero, Andrea, Joshua C. C. Chan, Todd E. Clark, and Massimiliano Marcellino (2022), "Corrigendum to: Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 227, 506–512.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019), "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212, 137–154.

Carriero, Andrea, Todd E. Clark, Massimiliano Marcellino, and Elmar Mertens (2024), "Addressing COVID-19 outliers in BVARs with stochastic volatility," *The Review of Economics and Statistics*, 106, 1403–1417.

Chernis, Tony, Niko Hauzenberger, Florian Huber, Gary Koop, and James Mitchell (2025), "Predictive density combination using Bayesian machine learning," *International Economic Review*, 66, 1287–1315.

Clark, Todd E., Gergely Ganics, and Elmar Mertens (2022), "What is the predictive value of SPF point and density forecasts?" Working paper 22-37, Federal Reserve Bank of Cleveland.

Clark, Todd E., and Elmar Mertens (2024), "Survey expectations and forecast uncertainty," in *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, Cheltenham, UK: Edward Elgar Publishing, 305–333.

Clements, Michael P. (2004), "Evaluating the Bank of England density forecasts of inflation," *Journal of Forecasting*, 23, 463–473.

——— (2008), "Survey-based expectations and the Phillips curve," *Journal of Macroeconomics*, 30, 555–572.

——— (2010), "Explanations of the inconsistencies in survey respondents' forecasts," *European Economic Review*, 54, 536–549.

——— (2014), "Probability distributions or point predictions? Survey forecasts of US output growth and inflation," *International Journal of Forecasting*, 30, 99–117.

——— (2018), "Are macroeconomic density forecasts informative?" *International Journal of Forecasting*, 34, 181–198.

——— (2025), "Inconsistent survey histograms and point forecasts revisited," *Journal of Economic Behavior & Organization*, 236.

Clements, Michael P., Robert W. Rich, and Joseph Tracy (2025), "An investigation into the uncertainty revision process of professional forecasters," *Journal of Economic Dynamics and Control*, 173.

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.

Coroneo, Laura, Fabrizio Iacone, and Fabio Profumo (2024), "Survey density forecast comparison in small samples," *International Journal of Forecasting*, 40, 1486–1504.

Cover, Thomas A., and Joy A. Thomas (2006), *Elements of Information Theory 2nd Edition*, Wiley Series in Telecommunications and Signal Processing: Wiley, 2nd edition.

Croushore, Dean, and Tom Stark (2001), "A real-time data set for macroeconomists," *Journal of Econometrics*, 105, 111–130.

Csiszár, Imre (1975), "I-divergence geometry of probability distributions and minimization problems," *Annals of Probability*, 3, 146–158.

Deming, W. Edwards, and Frederick F. Stephan (1940), "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known," *Annals of Mathematical Statistics*, 11, 427–444.

Diebold, Francis X., Minchul Shin, and Boyuan Zhang (2023), "On the aggregation of probability assessments: Regularized mixtures of predictive densities for Eurozone inflation and real interest rates," *Journal of Econometrics*, 237, p. 105321.

Engelberg, Joseph, Charles F. Manski, and Jared Williams (2009), "Comparing the point predictions and subjective probability distributions of professional forecasters," *Journal of Business & Economic Statistics*, 27, 30–41.

Faust, Jon, and Jonathan H. Wright (2013), "Forecasting inflation," in *Handbook of Economic Forecasting* eds. by Graham Elliott, and Allan Timmermann, 2A of Handbook of Economic Forecasting: Elsevier, chap. 1, 2–56.

Fienberg, Stephen E. (1970), "The iterative proportional fitting procedure," *Biometrika*, 57, 617–624.

Galvão, Ana Beatriz, Anthony Garratt, and James Mitchell (2021), "Does judgment improve macroeconomic density forecasts?" *International Journal of Forecasting*, 37, 1247–1260.

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.

Giacomini, Raffaella, and Giuseppe Ragusa (2014), "Theory-coherent forecasting," *Journal of Econometrics*, 182, 145–155.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (2004), "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185–212.

Kenny, Geoff, Thomas Kostka, and Federico Masera (2014), "How informative are the subjective density forecasts of macroeconomists?" *Journal of Forecasting*, 33, 163–185.

——— (2015), "Density characteristics and density forecast performance: A panel analysis," *Empirical Economics*, 48, 1203–1231.

Knüppel, Malte, and Lora Pavlova (2023), "Assessing the ex-ante uncertainty in the US SPF," Mimeo.

Krueger, 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.

Krueger, Fabian, and Lora Pavlova (2024), "Quantifying subjective uncertainty in survey expectations," *International Journal of Forecasting*, 40, 796–810.

Lenza, Michele, Inès Moutachaker, and Joan Paredes (2025), "Density forecasts of inflation: A quantile regression forest approach," *European Economic Review*, 178, p. 105079.

Meyler, Aidan (2020), "Forecast performance in the ecb spf: ability or chance?" ECB Working Paper Series 2371, European Central Bank.

Montes-Galdón, Carlos, Joan Paredes, and Elias Wolf (2026), "A robust approach to tilting: Parametric relative entropy," Working Paper Series 3200, European Central Bank.

Rich, Robert W., Joseph Song, and Joseph Tracy (2012), "The measurement and behavior of uncertainty: evidence from the ecb survey of professional forecasters," Staff Report 588, Federal Reserve Bank of New York.

Robertson, John C., Ellis W. Tallman, and Charles H. Whiteman (2005), "Forecasting using relative entropy," *Journal of Money, Credit and Banking*, 37, 383–401.

Schennach, Susanne M. (2007), "Point estimation with exponentially tilted empirical likelihood," *The Annals of Statistics*, 35, 634–672.

Stock, James H., and Mark W. Watson (2016), "Core inflation and trend inflation," *The Review of Economics and Statistics*, 98, 770–784.

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.

Tallman, Emily, and Mike West (2022), "On entropic tilting and predictive conditioning," Papers 2207.10013, arXiv.org.

Waggoner, Daniel F., and Tao Zha (1999), "Conditional forecasts in dynamic multivariate models," *The Review of Economics and Statistics*, 81, 639–651.

West, Mike (2024), "Perspectives on constrained forecasting," *Bayesian Analysis*, 19, 1013–1039.

Table 1: Accuracy of tilted and untilted forecasts from BVAR-SVO-t model: 1996-2025

	RMSE						CRPS					
h	Output	BFI	Unempl.	Payroll	Inflation	5y yield	Output	BFI	Unempl.	Payroll	Inflation	5y yield
0	0.51	0.78	0.34	0.62	0.85	0.98	0.73	0.90	0.55	0.68	0.92	0.99
1	0.71	0.89	0.75	0.65	0.87	0.98	0.75*	0.95	0.76	0.67	0.94	0.98
2	0.79	0.95	0.82	0.64	0.86	0.98	0.76*	0.95	0.83	0.69*	0.94	0.99
3	0.73*	0.92	0.74	0.61	0.83	0.96	0.75*	0.92	0.81	0.67**	0.93	0.96
4	0.87	0.93	0.69	0.75	0.83	0.94**	0.80	0.92	0.79	0.71**	0.93	0.94**
5	0.84	0.90	0.69	0.77**	0.79	0.94**	0.79	0.90	0.80	0.71**	0.90	0.93**
6	0.88	0.94	0.72	0.83**	0.77	0.93***	0.79	0.91	0.81	0.74**	0.88**	0.93**
7	0.85	0.92	0.75	0.86*	0.77	0.93***	0.79	0.88	0.82	0.76**	0.86**	0.94**
8	0.86	0.89	0.79	0.88*	0.76	0.93**	0.80	0.86	0.84	0.79*	0.86**	0.94*
9	0.86*	0.86	0.82	0.90	0.77	0.92**	0.82*	0.84	0.87	0.83	0.87**	0.93*
10	0.90	0.86*	0.85	0.93	0.78	0.92**	0.85	0.84	0.89	0.87	0.87**	0.92*
11	0.92	0.87*	0.87	0.95	0.79	0.92**	0.87	0.85*	0.91	0.91	0.88*	0.92*
12	0.94	0.89*	0.90	0.97	0.81	0.92**	0.90	0.87*	0.93	0.94	0.89*	0.92*
13	0.96	0.91*	0.91	0.98	0.82	0.92**	0.94	0.90*	0.94	0.97	0.90**	0.92*
14	0.99	0.95	0.94	1.00	0.82	0.92**	0.98	0.94	0.96	1.01	0.89**	0.92*
15	1.01	0.98	0.96	1.01	0.83	0.92**	1.01	0.98	0.98	1.03	0.89**	0.92*

Note: h quarters ahead forecasts. Relative RMSE and CRPS of BVAR-SVO(t)-AR(1) model BVAR-SVO(t)-AR(1) tilted against SPF histograms for inflation, unemployment and GDP, with BVAR-SVO(t)-AR(1) in denominator. Evaluation window from 1996Q1 through 2025Q1 (and as far as realized values and tilting targets are available). Reflecting availability of SPF data, the first forecast origin used in this evaluation is 1996Q1. 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 2: Accuracy of tilted and untilted forecasts from BVAR-SVO-t model: 2007-2025

	RMSE						CRPS					
h	Output	BFI	Unempl.	Payroll	Inflation	5y yield	Output	BFI	Unempl.	Payroll	Inflation	5y yield
0	0.47	0.73	0.33	0.62	0.81	0.99	0.65	0.85	0.48	0.65	0.88	1.00
1	0.68	0.87	0.74	0.64	0.83	0.99	0.66**	0.91	0.72	0.64	0.91	0.99
2	0.75	0.94	0.81	0.64	0.82	0.99	0.66**	0.90	0.80	0.66*	0.91	0.99
3	0.68**	0.88	0.73	0.60	0.78	0.95	0.62**	0.84	0.77	0.64*	0.89	0.95
4	0.82*	0.88	0.68	0.75	0.79	0.93**	0.69**	0.82*	0.74	0.68**	0.89	0.93*
5	0.79*	0.82**	0.68	0.76**	0.74	0.92**	0.70*	0.78*	0.74	0.68**	0.85	0.91**
6	0.85	0.87	0.70	0.83**	0.71	0.92***	0.72*	0.80	0.76	0.71**	0.82**	0.90**
7	0.83	0.87	0.73	0.86*	0.71	0.91***	0.72*	0.79	0.77	0.73*	0.80***	0.91**
8	0.84	0.84	0.77	0.87*	0.71	0.91**	0.74*	0.77	0.80	0.75*	0.80**	0.92*
9	0.83*	0.78*	0.80	0.89	0.72	0.90**	0.74*	0.74*	0.83	0.78	0.81**	0.90*
10	0.87	0.74**	0.83	0.92	0.72	0.90**	0.78*	0.73*	0.86	0.82	0.81**	0.89*
11	0.88*	0.70**	0.85	0.93	0.74	0.89**	0.79*	0.73*	0.88	0.85	0.83**	0.89*
12	0.90*	0.67**	0.87	0.94	0.76	0.89**	0.82**	0.73**	0.90	0.88	0.84**	0.89*
13	0.93**	0.72**	0.89	0.96*	0.77	0.89**	0.88**	0.78**	0.92	0.92	0.85**	0.89*
14	0.97*	0.82**	0.91	0.98	0.77	0.89**	0.95	0.86**	0.94	0.97	0.84**	0.89*
15	1.00	0.89	0.94	0.99	0.77	0.89**	1.00	0.93	0.97	1.00	0.84**	0.89*

Note: h quarters ahead forecasts. Relative RMSE and CRPS of BVAR-SVO(t)-AR(1) model BVAR-SVO(t)-AR(1) tilted against SPF histograms for inflation, unemployment and GDP, with BVAR-SVO(t)-AR(1) in denominator. Evaluation window from 2007Q1 through 2025Q1 (and as far as realized values and tilting targets are available). Reflecting availability of SPF data, the first forecast origin used in this evaluation is 2007Q1. 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 tilted to selected SPF percentiles: 1996-2025

	RMSE						CRPS					
h	Output	BFI	Unempl.	Payroll	Inflation	5y yield	Output	BFI	Unempl.	Payroll	Inflation	5y yield
Tilting to 50th percentiles of SPF histograms												
0	1.42*	1.09	0.72	0.59	0.90	1.32	1.16	1.09	0.89	0.70	0.98	1.09**
3	1.24	1.13	0.87	0.69	0.90	1.17*	0.98	1.02	0.96	0.82	0.98	1.05
7	0.83	0.87	0.84	0.86**	0.82	1.07	0.83*	0.89	0.88	0.83**	0.94	0.99
11	0.94	0.90	0.91	0.98	0.83	1.03	0.95	0.93	0.93	0.97	0.96	0.96
15	1.03	0.99	0.98	1.01	0.85	0.99	1.06*	1.03	0.99	1.05***	0.98	0.96
Tilting to 15th, 50th, and 85th percentiles of SPF histograms												
0	0.85	0.91	0.73	0.66	0.90	1.13**	1.13	1.09	1.01	0.87	0.99	1.12***
3	0.75	0.92	0.77	0.65	0.82	1.01	0.79	0.96	0.86	0.72	0.93	1.00
7	0.84	0.91	0.75	0.84**	0.76	0.97	0.78*	0.87	0.80	0.76**	0.86**	0.96*
11	0.93	0.88	0.88	0.96	0.79	0.95	0.89	0.87	0.90	0.92	0.89	0.94
15	1.02	0.97	0.96	1.00	0.82	0.94*	1.04	1.00	0.98	1.03	0.91*	0.93

Note: h quarters ahead forecasts. Relative RMSE and CRPS of BVAR-SVO(t)-AR(1) BVAR-SVO(t)-AR(1) tilted to inflation, unemployment and GDP (untilted BVAR-SVO(t)-AR(1) in denominator). Evaluation window from 1996Q1 through 2025Q1 (and as far as realized values and tilting targets are available). Reflecting availability of SPF data, the first forecast origin used in this evaluation is 1996Q1. 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: Forecast accuracy when tilting to different targets (1996-2025)

	RMSE					CRPS				
					Output					
h	Baseline	0+1y	GDP	U	Inf	Baseline	0+1y	GDP	U	Inf
0	0.51	0.51	0.44	0.55	0.98	0.73	0.73	0.69	0.77	1.00
3	0.73*	0.74	0.70*	0.73	1.00	0.75*	0.75*	0.74*	0.76**	1.00
7	0.85	0.94	0.83	0.86	0.98***	0.79	0.90	0.77*	0.80	0.96**
11	0.92	1.02	0.92	0.91**	0.96***	0.87	1.02	0.86	0.89***	0.95***
15	1.01	1.05	0.99	1.06	0.98*	1.01	1.06	0.98	1.12	0.97**
					BFI					
0	0.78	0.78	0.74	0.90	0.99	0.90	0.89	0.87	0.98	1.00
3	0.92	0.94	0.81	0.90	1.00	0.92	0.93	0.88	0.90	1.00
7	0.92	0.96	0.88	0.86	0.99***	0.88	0.93	0.86	0.76	0.98***
11	0.87*	0.99	0.88*	0.71**	0.94***	0.85*	1.00	0.86*	0.77**	0.95***
15	0.98	1.04	0.97	1.02	0.95**	0.98	1.06	0.96	1.07	0.96**
					Unempl.					
0	0.34	0.34	0.89	0.30	1.00	0.55	0.56	0.91	0.45	1.00
3	0.74	0.75	0.85	0.71	1.00	0.81	0.82	0.87*	0.76	1.00
7	0.75	0.74	0.86*	0.66	1.01	0.82	0.82	0.88	0.72	1.01
11	0.87	0.87	0.92	0.77	1.01	0.91	0.90	0.91	0.83	1.00
15	0.96	0.96	0.98	0.89	1.01	0.98	0.97	0.98	0.92	1.01
					Payroll					
0	0.62	0.62	0.83	0.56	0.99	0.68	0.68	0.83	0.53	1.00
3	0.61	0.61	0.77	0.60	0.98	0.67**	0.67**	0.74**	0.63	1.01
7	0.86*	0.91	0.86**	0.87	1.01	0.76**	0.83*	0.79**	0.73*	1.02
11	0.95	1.02	0.96	0.92	1.00	0.91	1.02	0.92	0.83*	1.00
15	1.01	1.04	1.01	1.00	1.00	1.03	1.08	1.02	1.01	1.00
					Inflation					
0	0.85	0.85	0.93	0.90	0.82	0.92	0.92	0.96	0.90	0.90
3	0.83	0.82	0.95	0.92	0.77	0.93	0.93	0.99	0.97	0.89
7	0.77	0.76	0.93	0.88	0.73	0.86**	0.86**	0.98	0.97	0.81**
11	0.79	0.77	0.91	0.88	0.79	0.88*	0.87*	0.97	1.00	0.86*
15	0.83	0.83	0.93	0.88	0.87	0.89**	0.90**	0.98	0.99	0.91
					5y yield					
0	0.98	0.98	0.96**	1.13	1.02	0.99	0.98	0.96*	1.01	1.03
3	0.96	0.96*	0.95***	1.00	1.07**	0.96	0.96	0.94***	0.95	1.09**
7	0.93***	0.94**	0.95***	0.92**	1.06*	0.94**	0.95	0.94***	0.89**	1.10**
11	0.92**	0.95*	0.95***	0.91	1.05	0.92*	0.96	0.93**	0.86	1.10*
15	0.92**	0.95***	0.95***	0.88*	1.04	0.92*	0.97**	0.94**	0.83*	1.08*

Note: h quarters ahead forecasts. Relative RMSE and CRPS of BVAR-SVO(t)-AR(1), tilted against different SPF histogram targets, with BVAR-SVO(t)-AR(1) in denominator. Sample: "full sample". Tilting targets as indicated in column headers. 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.

Appendices

A Sufficient condition in schematic example of Figure 1

In the schematic example of Figure 1, the proposal density lacks mass for the joint case of low GDP growth and high inflation, which violates the sufficient condition of Assumption 3, while still satisfying the necessary condition of Assumption 1. A solution to the tilting problem may still exist, but only if it is feasible to construct a proper density q that satisfies not only the marginal constraints from the SPF but also the added zero restriction inherited from the proposal. Formally, denote the joint probabilities of q as q_{11}, q_{12}, q_{21}, and q_{22}, where the first index corresponds to GDP growth (low or high) and the second index corresponds to inflation (low or high). The marginal probabilities are denoted as $q_1^{\text{GDP}} = q_{11} + q_{12}$ for low GDP growth, $q_2^{\text{GDP}} = q_{21} + q_{22}$ for high GDP growth, $q_{\cdot 1}^{\text{INF}} = q_{11} + q_{21}$ for low inflation, and $q_{\cdot 2}^{\text{INF}} = q_{12} + q_{22}$ for high inflation.

We take it as given that the SPF probabilities sum to one, so that $q_{1\cdot}^{\text{GDP}} + q_{2\cdot}^{\text{GDP}} = 1$ and $q_{\cdot 1}^{\text{INF}} + q_{\cdot 2}^{\text{INF}} = 1$, so that the SPF imposes two linear restrictions on the four entries of q. The zero restriction from the proposal, $q_{12} = 0$, then implies $q_{11} = q_{1\cdot}^{\text{GDP}}$ and $q_{22} = q_{\cdot 2}^{\text{INF}}$. With the additional restriction that all entries of q sum to one, the remaining entry is pinned down as well, leaving only the non-negativity requirement to be satisfied for a proper density to exist:

$$q_{21} = 1 - q_{11} - q_{22} = 1 - q_{1\cdot}^{\text{GDP}} - q_{\cdot 2}^{\text{INF}} \geq 0 \quad \iff \quad q_{1\cdot}^{\text{GDP}} + q_{\cdot 2}^{\text{INF}} \leq 1\,. \tag{42}$$

As a result, a feasible solution exists only if the sum of the SPF's marginal probabilities for low growth and high inflation do not exceed one.

B Details of the computational implementation

[TO BE WRITTEN]

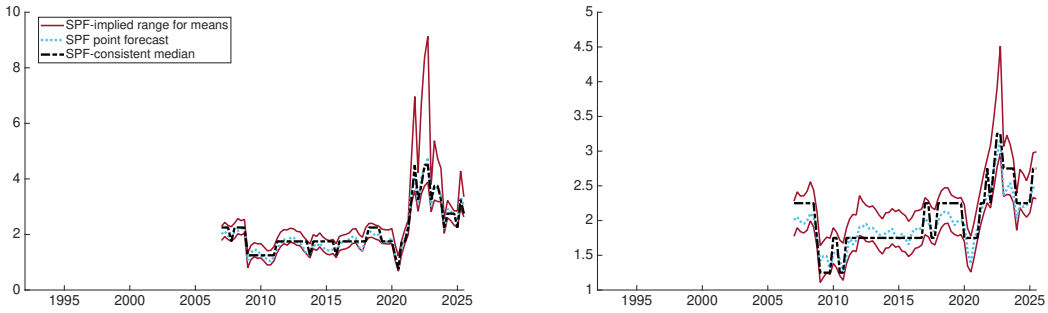
C Additional figures

This Appendix provides additional figures. Figures 13 and 14 extend the results on coherence between SPF point and histogram, shown in Figure 4 above for the case of GDP growth, to SPF histograms for core inflation and the unemployment rate.

Figure 15 provides a summary of the number of iterations required for the iterative proportional fitting algorithm to converge for different sets of histogram targets across different forecast origins.

Figure 16 provides a summary of the percentage point mass of the tilted density $q(\cdot)$ for different sets of histogram targets that is assigned to cross-histogram events with zero mass under the original proposal density $p^0(\cdot)$. This mass represents probability redistributed to events that lacked support in the baseline BVAR-SVO-t forecast but receive positive mass in the tilted density after tilting to match SPF histograms. Apart from the onset of COVID-19, the mass assigned to zero-mass events rarely exceeds 10%, and the large amounts of mass (of over 60%) assigned in early 2020 is driven by the need to account for unemployment rate histograms.

Figure 13: SPF-implied mean and median forecasts for core PCE inflation

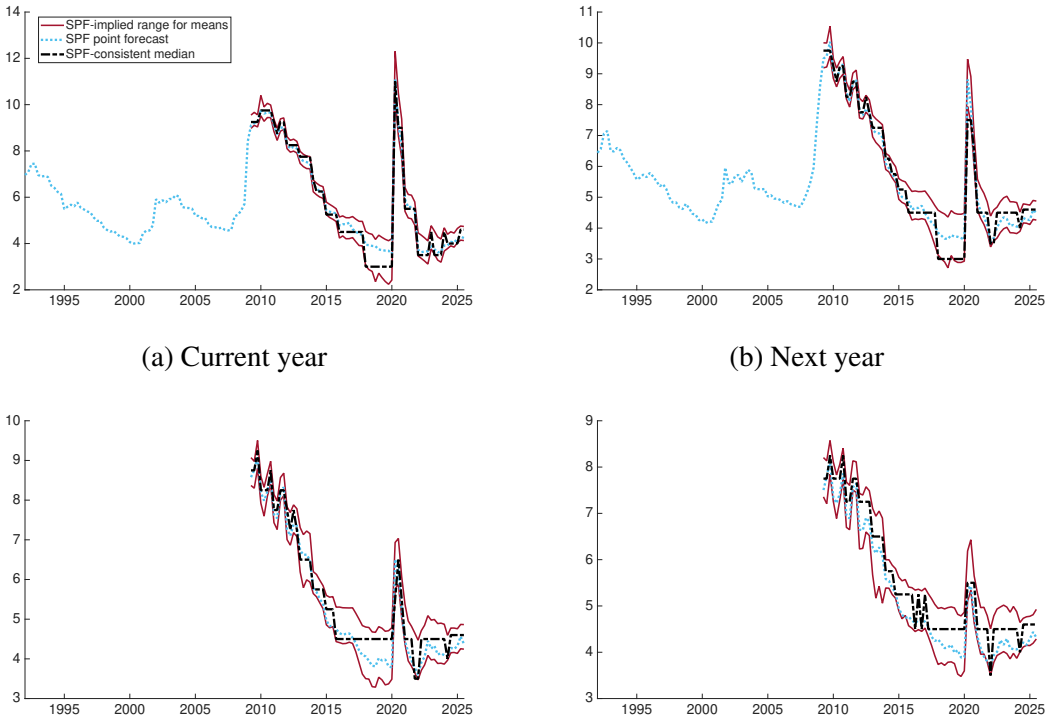


(a) Current year

(b) Next year

Note: SPF-consistent median is the midpoint of the bin containing the median. Upper and lower bounds for SPF-consistent means are computed from lower and upper bounds of the bins. The point forecasts are the average of individual forecasters' point forecasts. Core PCE inflation histograms are available from the SPF since 2007.

Figure 14: SPF-implied mean and median forecasts for unemployment rate



(a) Current year

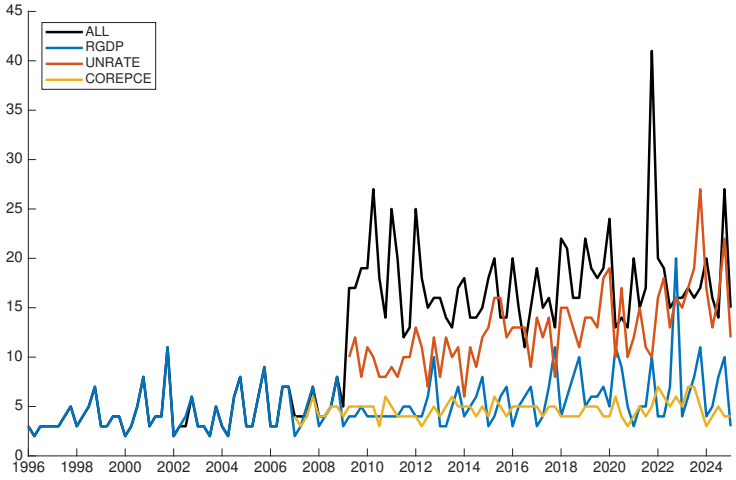
(b) Next year

(c) Two years ahead

(d) Three years ahead

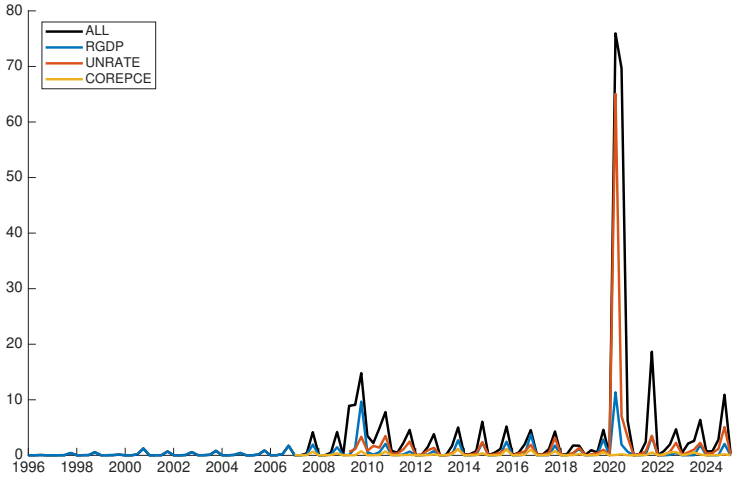
Note: SPF-consistent median is the midpoint of the bin containing the median. Upper and lower bounds for SPF-consistent means are computed from lower and upper bounds of the bins. The point forecasts are the average of individual forecasters' point forecasts. Unemployment rate histograms are available from the SPF since 2009.

Figure 15: Steps used in iterative proportional fitting algorithm



Note: The figure shows the number of iterations required for the iterative proportional fitting algorithm to converge for different sets of histogram targets across different forecast origins.

Figure 16: Tilted density mass assigned to zero-mass cross-histogram events



Note: The figure shows the percentage point mass of the tilted density $q(\cdot)$ for different sets of histogram targets that is assigned to cross-histogram events with zero mass under the original proposal density $p^0(\cdot)$. This mass represents probability redistributed to events that lacked support in the baseline BVAR-SVO-t forecast but receive positive probability through the augmented proposal density used in the entropic tilting procedure.