



EUROPEAN CENTRAL BANK
EUROSYSTEM

Working Paper Series

Todd E. Clark, Elmar Mertens

Entropic tilting of forecasts to SPF
histograms: analytics & applications

No 3284 / September 2026

Abstract

We develop a direct approach to incorporating survey density forecasts into model-based predictive distributions. Histogram forecasts from the U.S. Survey of Professional Forecasters (SPF) carry rich nonparametric information about expected outcomes, but existing methods rely on moment-based approximations that discard part of it. We instead tilt entropically to the histogram probabilities themselves, matching them exactly. After reformulating the single-histogram problem, we derive a new analytic characterization of the multiple-histogram case, solved by Iterative Proportional Fitting and applicable to simulated densities from essentially any model. Applying the method to real-time forecasts from a Bayesian VAR with time-varying volatility, we find that tilting to SPF histograms substantially improves accuracy relative to the model's baseline forecasts, especially during the Great Recession and the COVID-19 pandemic. The gains extend beyond the variables the SPF targets, improving forecasts for other variables in the system as well.

Keywords: Survey expectations, predictive densities, Bayesian vector autoregression, relative entropy, iterative proportional fitting

JEL Classification: C11, C53, E37

Non-technical summary

Central banks and other forecasters routinely rely on a combination of judgment and statistical models. An important, and publicly available, source of judgment is surveys of professional forecasters. Such surveys — conducted quarterly for the United States by the Federal Reserve Bank of Philadelphia and, in a similar format, for the euro area by the ECB — ask respondents not only for their best guess of outcomes such as GDP growth, unemployment, and inflation, but also for the probabilities they attach to different ranges of outcomes. These probability assessments, reported as histograms (a set of predefined outcome ranges, or “bins,” with a probability assigned to each), contain rich information about the uncertainty and risks surrounding the outlook — information that a single point forecast cannot convey. Yet most existing methods for feeding survey information into forecasting models use only the point forecasts, or compress the histograms into one or two summary statistics such as a mean and a variance, discarding much of their content. This paper develops a method that connects the full survey histograms with a statistical model, and asks whether doing so improves forecast accuracy.

The method works as follows. A forecasting model — in our application, a Bayesian vector autoregression (BVAR), a widely used statistical model that captures the joint dynamics of several macroeconomic variables — produces a large number of simulated future paths for the economy. Each path is a coherent scenario for GDP growth, unemployment, inflation, and other variables over the coming years. Our procedure then reweights these scenarios — giving more weight to some and less to others — so that the reweighted set of scenarios reproduces exactly the probabilities reported in the survey histograms, while otherwise staying as close as possible to the model’s original assessment. This reweighting method is known as entropic tilting, and “as close as possible” is measured by a statistical criterion of divergence between probability distributions. Because the adjustment only reweights complete scenarios rather than altering them, the economic relationships embedded in the model — for instance, that weak GDP growth tends to come with rising unemployment — are preserved.

The paper’s methodological contribution is to show that this reweighting problem has a simple and transparent solution when the targets are histogram probabilities. With a single histogram, the solution amounts to scaling up or down the weights of all scenarios that fall in the same bin, by the ratio of the survey probability to the model probability for that bin — leaving the distribution of scenarios within each bin untouched. With several histograms at once (for example, for multiple variables and multiple years), the survey pins down the probabilities for each histogram separately but says nothing about how they interact; we show that the solution then takes the form of a product of histogram-specific adjustment factors, which can be computed quickly and reliably by a classical algorithm known as iterative proportional fitting. The paper also resolves practical complications that arise when the model’s simulations place no weight on some combinations of survey outcomes. The resulting procedure is fast, exact, and applicable to simulated forecasts from essentially any model, not just the BVAR used here.

The empirical application uses data for the United States. At each quarterly forecast origin from 1996 through 2025, the model is estimated on the data as they were available at the time, and its forecasts are tilted to all survey histograms then published — for GDP growth, the unemployment rate, and core

inflation, at horizons from the current year up to three years ahead. Forecast accuracy is then evaluated against subsequently realized outcomes.

Three findings stand out. First, tilting to the survey histograms consistently improves the model's forecast accuracy, for both point forecasts and the full distribution of outcomes. Second, the gains extend beyond the variables targeted by the survey: forecasts of employment, investment, and a longer-term interest rate also improve, because the reweighting operates on complete economic scenarios and thus transmits the survey's information through the model's internal linkages. Third, the gains are concentrated in — though not confined to — periods of economic turbulence, notably the Great Recession and the COVID-19 pandemic, when professional forecasters' judgment adjusted more rapidly than the statistical model. Finally, tilting directly to the survey histograms performs about as well as the more common alternative of first fitting distributions to the histograms and tilting to the moments of those fitted distributions — while being computationally lighter, avoiding assumptions about the shape of the distribution within each histogram, and sidestepping the occasional difficulty of matching fitted moments exactly.

For central banks, the practical implication is that survey-based probability forecasts are a valuable input for model-based forecasting and risk assessment, and that their full information content can be exploited with a transparent, computationally light procedure. The method can directly be applied to the ECB's own Survey of Professional Forecasters, which collects probability histograms of the same format for euro-area growth, inflation, and unemployment, and can be used wherever model forecasts need to be reconciled with external probabilistic assessments — including, in future applications, survey-based recession probabilities.

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. Beyond point forecasts, the SPF (both in the U.S. and euro area) also provides density forecasts in the form of probability histograms over forecast bins, containing rich probabilistic information about forecast uncertainty and tail risks. 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. However, we document that some of these issues are more prevalent in individual forecasters' histograms than in the aggregate histograms that average across forecasters.

One way survey forecasts have been used is to improve forecasts from time series models, such as vector autoregressions, by combining alternative projections. More broadly, these approaches can be viewed as methods for integrating externally observed predictive information into model-based forecasting systems. 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. Robertson, Tallman, and Whiteman (2005) introduced into economic forecasting the closely related, less parametric approach of entropic tilting, which minimizes the relative entropy — also known as the Kullback–Leibler (KL) divergence or the Kullback–Leibler Information Criterion (KLIC) — between the original and modified distributions. 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.

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, Krüger, 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.¹ The existing work that uses survey density forecasts — from the U.S. SPF, the ECB's SPF for the euro area, or comparable surveys for other economies — to inform model-based forecasts does so by extracting a limited set of moments from the survey-based 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.² Ganics and Odendahl (2021) impute first and second moments

¹ Additional applications of entropic tilting to BVAR forecasts include Cogley, Morozov, and Sargent (2005) and Bobeica and Hartwig (2023).

² Specifically, Galvão, Garratt, and Mitchell (2021) extract moment parameters from fitted generalized beta distributions

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 direct approach to incorporating information from SPF density forecasts into model-based predictive systems. Our approach takes the SPF histograms as given — i.e., as primitive probabilistic objects — and tilts the predictive distribution of the BVAR to directly match SPF histogram probabilities. In doing so, we preserve the probabilistic information structure contained in the survey histograms without taking a stand on how the SPF forecaster distributes predictive mass within each histogram bin or imposing parametric assumptions on the survey side.³ In contrast, some earlier research tilted model forecasts to estimated moments fitted to SPF histograms, taking a stand on a parametric or non-parametric distribution to characterize the histogram and then in turn estimating the distribution. Throughout, we focus our application on histograms of the average SPF forecaster, also known as “consensus” histograms.⁴

Our paper has three primary contributions. The first contribution 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 reweights 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. Our novel implementation exploits sparsity in the BVAR’s proposal density, which proves crucial to computation with multiple histograms. Our implementation also addresses the challenge that, in practice, the set of draws from the model-based density may not have mass in all of the joint cross-histogram events. To address this case, we employ an augmented density that has mass in all joint cross-histogram events, obtained as a mixture of the original model-based density with a uniform density assigned a very low weight. This approach preserves the computational benefits of the proposal’s sparsity, since the uniform component factorizes across histograms.

Our second contribution is applied: Using our analytical solution for the tilting problem, we entropically tilt the quarterly forecast distributions of a BVAR 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. To the best

(Engelberg, Manski, and Williams, 2009), whereas Banbura, et al. (2024) employ random sampling from multinomial distributions and kernel-smoothed estimates from the resulting densities.

³On the model side, our empirical implementation employs auxiliary conventions to populate cross-histogram events for which the finite BVAR simulation delivers no draws, as described in Section 3.5.2 and Appendix E. These conventions matter only to the extent that such events receive noticeable tilted mass, which occurs chiefly at the onset of the COVID-19 pandemic.

⁴For a discussion of potential difficulties in handling individual SPF histograms see, for example, our survey in Clark and Mertens (2024).

⁵Our 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.

of our knowledge, our paper is the first to try to improve time series model forecasts for U.S. data using SPF histogram forecasts. By contrast, other studies have used point forecasts for the U.S. or histogram forecasts from other surveys for other economies. Compared to the latter, we consider tilting to a substantially larger set of histograms (as many as 10 for different variables and horizons, compared to about four in other studies), and we parse the effects of tilting to specific variables or horizons. Accordingly, our paper provides the first comprehensive examination of the potential to improve time series model forecasts with predictive content from SPF histograms for the U.S.

Our third contribution is empirical: We examine how information contained in SPF histograms propagates through multivariate predictive systems and shapes forecast accuracy for major macroeconomic aggregates. We provide evidence that, relative to the baseline model forecasts, the information content added by the SPF histograms has varied considerably over time, with most of the information coming from the SPF's GDP growth and unemployment rate forecasts. 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, by amounts that vary by horizon and across variables. Importantly, these gains extend beyond the model variables for which SPF histogram targets are available, also improving forecasts for other variables in the system (such as employment), indicating that survey density information propagates through the joint predictive structure of the multivariate model.

These average gains mask considerable variation over time. The benefits of entropically 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 publication of core PCE inflation forecasts only in 2007. 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 its aftermath and around COVID-19's outbreak in 2020 — the same episodes in which the KLIC distance indicates that the SPF histograms carry the most information.⁶ A comparison of our direct histogram tilting to the alternative of tilting BVAR forecasts to moments fitted to SPF histograms shows the approaches to yield similar accuracy of forecasts. Direct histogram tilting has advantages in computation time, circumventing occasional problems in fully matching the fitted moments, and fully incorporating the information in the SPF histograms.

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, and Section 6 concludes. Additional details and results are provided in a supplementary online appendix.

⁶As 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 — in particular, a standard Gaussian BVAR — are, of course, larger. 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.

2 Related literature

Broadly, our paper contributes to the literature evaluating the predictive content of probabilistic survey forecasts. A large part of this literature has examined heterogeneity, consistency, and performance characteristics across individual respondents.⁷ By contrast, we investigate the predictive value of aggregated probability forecasts, where issues of coherence and consistency are fairly negligible.

Our framework provides a method for integrating survey-implied predictive densities into multivariate forecasting systems. Methodologically, our use of entropic tilting to align model-implied forecast densities with survey histograms across multiple forecast horizons can be viewed as a modern and applied extension of the classical *Iterative Proportional Fitting* (IPF) algorithm (Deming and Stephan, 1940). For estimating cell probabilities in a contingency table, Ireland and Kullback (1968) recast IPF in information-theoretic terms, showing that its fixed point minimizes the discrimination information (relative entropy) from the initial table subject to the marginal constraints, that the optimal cell probabilities take a multiplicative cross-product form, and that the iteration converges; they also state the generalization to multi-way tables. Fienberg (1970) subsequently gave a complementary geometric proof of convergence, and Csiszár (1975) embedded the procedure in the general theory of I-projections. 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 KL divergence from a prior distribution. Our framework departs from this classical literature in several respects—most importantly, our baseline is a continuous predictive density rather than a finite table of cell probabilities estimated from data, and our constraints come from external survey histograms over distinct, generally nonlinear transformations of the outcome vector. These differences raise issues of interaction structure and sparse support in forms specific to our setting — cross-histogram events generated by nonlinear transformations of a continuous predictive density, with support determined by a finite simulation — that the classical treatments, which start from an observed contingency table, do not have to confront. Because making these contributions precise requires the machinery developed below, we defer their detailed discussion to Section 3.6, after we have introduced the relevant concepts. Our approach also connects 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 Kullback–Leibler-projection principle to impose nonlinear theoretical moment conditions (e.g., Euler equation restrictions) on model-based forecast densities, establishing that entropic tilting improves forecast accuracy 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 in structural VARs can likewise be derived as KL-minimizing

⁷See, 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), Coroneo, Iacone, and Profumo (2024), Allayioti, et al. (2024), Clements, Rich, and Tracy (2025), and Clements (2025).

projections. In a different vein, Lewis and Whiteman (2015) apply entropic tilting to the posterior distribution of VAR parameters, reweighting draws to minimize out-of-sample forecast error subject to a KL divergence penalty.

Our focus on directly tilting model-based forecasts to match histogram probabilities provides a practical method to leverage the information content of survey histograms without imposing additional parametric assumptions.⁸ Our evaluation therefore differs from Giacomini and Ragusa (2014), whose frequentist rolling-window exercise tests whether theory-coherent moment restrictions improve forecast performance. By contrast, we evaluate empirical coherence: tilting Bayesian predictive densities to survey-implied distributions from the SPF and assessing calibration gains using standard point and density forecast accuracy metrics.

Our methodology builds directly on recent developments in entropic tilting, most notably Tallman and West (2022) and West (2024), who develop entropic tilting as a Bayesian decision-analytic method for modifying forecast distributions under expectation constraints, including quantile- and indicator-function targets. For the single-histogram case, we reformulate their solution in terms of multiplicative tilting weights that are constant within histogram bins. This reformulation makes explicit a structural property that is not immediately apparent from the general formulation: probability mass is redistributed only across histogram bins while preserving the within-bin shape of the original predictive density. Although mathematically equivalent to the original formulation, this characterization provides the key insight that allows us to develop analytical results for multiple histograms and the associated iterative proportional fitting (IPF) algorithm. The reformulation is not merely expositional. It reveals the multiplicative structure of histogram tilting that underlies our characterization of interaction entropy, the cross-product representation of the optimal solution, and the computationally efficient iterative proportional fitting algorithm developed below.

Recent work by Adrian, et al. (2025, 2026b,a) also builds on the entropic-tilting tradition in macroeconomic forecasting. Their *Scenario Synthesis* framework uses entropic tilting to turn partially specified scenarios—each summarized by a point forecast or a set of percentiles—into full predictive densities, before weighting and combining these through Bayesian predictive synthesis. On the tilting step itself, our approaches are close cousins: both project a baseline density onto distributional constraints in the Kullback–Leibler sense, in the shared lineage of Robertson, Tallman, and Whiteman (2005) and Tallman and West (2022). The differences lie elsewhere. Our targets are full histogram bin probabilities for several variables across multiple forecast horizons, imposed jointly rather than as selected percentiles or modes of an individual scenario, and reconciling these joint targets is what gives rise to the interaction-entropy decomposition and IPF algorithm developed below. The expected misclassification rate (EMR), which carries weight alongside entropic tilting already in Adrian, et al. (2025) and becomes the central

⁸Knüppel and Pavlova (2026) document that changes in SPF histogram bin widths induce structural breaks in measured aggregate uncertainty when uncertainty is computed by fitting smooth distributions to individual histograms. Our approach is unaffected: we tilt directly to bin probabilities and inherit within-bin densities from the BVAR, so no fitted variance enters our procedure. A related concern is that bin probabilities themselves may reflect bin-width-dependent reporting behavior (Glas and Hartmann, 2022); within our sample, only RGDP (2020Q2, 2024Q2) and UNRATE bin definitions changed materially, while COREPCE bins have been stable since 2007Q1.

object in Adrian, et al. (2026a), governs a different step: how multiple and possibly competing narrative scenarios are weighted against a reference distribution. That problem does not arise in our setting, where the SPF furnishes a single forecast (with multiple histograms) to be assimilated rather than a set of scenarios to be scored.

Seeking to address robustness concerns that can arise when moment-based targets lie far from the support of the baseline density (e.g., as illustrated by Banbura, et al. (2024) during the COVID-19 pandemic), Montes-Galdón, Paredes, and Wolf (2026) propose “parametric tilting” as an alternative to entropic tilting. Rather than directly reweighting draws, their approach minimizes the KL divergence between the model density and a skew- t distribution subject to moment constraints (e.g., mean, variance, or skewness) on the target distribution. The resulting density inherits the parametric regularity of the skew- t family, guaranteeing unimodality and numerical stability even when target moments differ substantially from those of the baseline. Our approach is complementary: Rather than fitting a parametric distribution to survey information, we treat the published SPF histogram probabilities as primitive objects and impose them directly as probabilistic constraints on the model-based predictive density. In contrast, the approach of Montes-Galdón, Paredes, and Wolf (2026) restricts the tilted density to a parametric skew- t family, whose parameters are chosen to minimize the KL divergence from the model density subject to survey moment constraints. More broadly, the role of the skew- t family distinguishes three approaches to incorporating survey information: Montes-Galdón, Paredes, and Wolf (2026) obtain their tilted density as the KL-closest member of a skew- t family subject to survey moment constraints (rather than matching a pre-specified skew- t density); we tilt toward nonparametric, bin-exact histogram targets; and Adrian, et al. (2026b), when they take the SPF as a reference distribution, fit a skew- t to survey-implied quantiles only as a *smoothing device* for that reference, with nothing tilted toward it.⁹

3 Analytics of tilting to histograms

Here, we present our new analytical solution to the problem of tilting to histograms. We begin by 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.

⁹The two uses of the skew- t are distinct. In Montes-Galdón, Paredes, and Wolf (2026) it is the parametric family within which the KL tilt of the model density is computed, proposed explicitly as a more stable alternative to standard entropic tilting. In Adrian, et al. (2026b) it merely turns a handful of survey quantiles into a continuous reference density; the survey-matching criterion applied to that reference is EMR, not KL. This split also locates each approach relative to the bin-width sensitivities noted by Knüppel and Pavlova (2026): those concerns attach to moment extraction from histograms, so they bear on the parametric target of Montes-Galdón, Paredes, and Wolf (2026), are largely moot for our bin-exact targets, and are muted for Adrian, et al. (2026b), whose skew- t enters only as a fixed reference rather than an evaluated forecast.

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 model-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 directly on draws generated from the model’s predictive density (without regard for its underlying structure), it can also be applied to models other than a BVAR, such as (nonlinear) DSGE models, quantile VARs, etc.

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 predictive density for outcomes at horizons $h = 0, 1, \dots, H$. The horizon $h = 0$ refers to the nowcast for y_t , which is typically unknown in real time at t . We denote the vector of outcomes at these horizons as $\mathbf{X}_t = (y_t, y_{t+1}, \dots, y_{t+H})$.

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 nonlinear transformations of $\mathbf{X}_t$. We consider jointly tilting to N SPF histograms indexed by $n = 1, 2, \dots, N$. For SPF histogram n , we denote the applicable transformation at origin t by $f_t^n(\mathbf{X}_t)$. 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 are time-dependent and may involve multiple (though typically not all) elements of $\mathbf{X}_t$. For example, for the unemployment rate, $f_t^n(\mathbf{X}_t)$ is the linear average of the quarterly unemployment rates for the targeted year contained in $\mathbf{X}_t$. However, for GDP growth and (core) PCE inflation the corresponding definitions of $f_t^n(\mathbf{X}_t)$ are nonlinear transformations of $\mathbf{X}_t$. Specifically, for GDP growth, $f_t^n(\mathbf{X}_t)$ 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 core PCE inflation, $f_t^n(\mathbf{X}_t)$ corresponds to the Q4 over Q4 change in the level of the core PCE price index. For growth and inflation, the fact that $\mathbf{X}_t$ consists of differences in log levels (rather than simple growth rates) adds another source of nonlinearity in the definitions of $f_t^n(\mathbf{X}_t)$.

Let y_t^{GDP} , y_t^{UNRATE} , and y_t^{INF} denote the elements of y_t that correspond to the quarterly values of annualized log-differenced real GDP, the unemployment rate, and annualized 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:

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

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

$$\text{Inflation: } f_t^n(\mathbf{X}_t) = 100 \times \exp \left(\sum_{i=h-3}^h y_{t+i}^{\text{INF}} / 400 \right) - 100 \quad (3)$$

where $Y_{t+k}^{\text{GDP}} \equiv \exp \left(\sum_{i=h-6}^k y_{t+i}^{\text{GDP}} / 400 \right)$ denotes the quarterly level of GDP in $t+k$ relative to the base period $t+h-7$. For $i < 0$ the values of y_{t+i}^{UNRATE} , y_{t+i}^{GDP} , and y_{t+i}^{INF} reflect fixed and known data points, whereas for $i \geq 0$ their values are generated by the model. Transformations targeting outcomes in other calendar years are defined analogously.

As discussed in more detail 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

$$\widehat{\text{Prob}}_t (b_{t,j-1}^n < f_t^n(\mathbf{X}_t) \leq b_{t,j}^n) \forall j, n, \quad (4)$$

where $b_{t,j}^n$ denotes the upper limit for the j th 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 :

$$A_{t,j}^n \equiv \{\mathbf{X}_t : b_{t,j-1}^n < f_t^n(\mathbf{X}_t) \leq b_{t,j}^n\}, \quad (5)$$

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

Henceforth, for simplicity we drop the t subscripts on the set of outcomes, bin limits, $\mathbf{X}$, and the transformation function f^n . Each histogram provides probabilities for a set of J mutually exclusive and exhaustive events $\{A_j^n\}_{j=1}^J$. For the baseline density, $p(A_j^n) \equiv p(\mathbf{X} \in A_j^n)$ describes the probabilities for $\mathbf{X}$ to fall in the j th 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 $q(\mathbf{X})$ is required to match the histogram probabilities, so that $q(A_j^n) \equiv q(\mathbf{X} \in A_j^n)$ 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, $q(\mathbf{X})$, that minimizes the relative entropy (or Kullback–Leibler divergence measure)

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

¹⁰For simplicity, the notation in this section treats all histograms as having an identical number of J bins. More generally, our implementation recognizes that the number of bins can vary across histograms, and denotes the number of bins of histogram n by $J(n)$.

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

$$\min_q D(q(\mathbf{X}) \parallel p(\mathbf{X})) \quad \text{s.t.} \quad q(\mathbf{X} \in A_i^n) = \widehat{\text{Prob}}(\mathbf{X} \in A_i^n) \quad \forall i, n. \quad (7)$$

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 $q(\mathbf{X} \in A_i^n) = E(\mathbb{1}(\mathbf{X} \in A_i^n))$ are particular moments of the distribution $q(\mathbf{X})$, 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 an 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 leveraging 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.

Since relative entropy involves the ratio of the tilted density to the original predictive density, $q(\mathbf{X})/p(\mathbf{X})$, the solution to the tilting problem is only well-defined if $p(\mathbf{X}) > 0$ for all $\mathbf{X}$ for which $q(\mathbf{X}) > 0$. A necessary condition for a solution to exist is that the support of $p(\mathbf{X})$ 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 $\widehat{\text{Prob}}(\mathbf{X} \in A_i^n) > 0$, we require that $p(\mathbf{X} \in A_i^n) > 0$.*

In our application, violations of Assumption 1 are limited to a few unemployment rate histograms and are quantitatively almost negligible, in that the SPF assigns no more than 0.1% probability to bins where the proposal density has zero mass.

A sufficient condition for a solution to exist, provided the target histograms are jointly feasible, is that the support of $p(\mathbf{X})$ is the entire space of $\mathbf{X}$, so that $p(\mathbf{X}) > 0$ for all $\mathbf{X}$.

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

As argued further below, in Section 3.3, 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 3.5 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.¹¹

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 $q(\mathbf{X})$ that matches a given set of marginal distributions, $\{q(A_j)\}_{j=1}^J$, and otherwise minimizes the relative entropy $D_x(q || p)$. The densities $q(\mathbf{X})$ and $p(\mathbf{X})$ can generally be factorized as

$$q(\mathbf{X}) = q(A_j)q(\mathbf{X}|A_j), \quad p(\mathbf{X}) = p(A_j)p(\mathbf{X}|A_j), \quad \forall j, \mathbf{X} \in A_j, \quad (8)$$

leading to the following decomposition of relative entropy:

$$\begin{aligned} D(q(\mathbf{X}) || p(\mathbf{X})) &= D(q(A) || p(A)) \\ &\quad + \sum_j q(A_j) \cdot \{D(q(\mathbf{X}|A_j) || p(\mathbf{X}|A_j))\}, \\ \text{where } D(q(A) || p(A)) &= \sum_j q(A_j) \cdot \log \left(\frac{q(A_j)}{p(A_j)} \right), \\ \text{and } D(q(\mathbf{X}|A_j) || p(\mathbf{X}|A_j)) &= \int_{x \in A_j} q(\mathbf{X}|A_j) \cdot \log \left(\frac{q(\mathbf{X}|A_j)}{p(\mathbf{X}|A_j)} \right) dx. \end{aligned}$$

Notice that $D(q(A) || p(A))$ 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 the baseline and the SPF targets is also the *minimized* relative entropy between the baseline density $p(\mathbf{X})$ and its tilted counterpart $q(\mathbf{X})$. To see this, note that we can freely choose the *conditionals* of the tilted density $q(\mathbf{X}|A)$, 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:

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

Because relative entropy is nonnegative, this is the minimizing solution to the tilting problem. Intuitively speaking, the minimizing solution reweights draws of $\mathbf{X}$ 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 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

¹¹In terms of notation, the original model-based proposal will be denoted p^0 and the augmented density will be denoted p , so that the tilting formulas described can continue to apply to a density that satisfies necessary and sufficient conditions for a solution to the tilting problem.

other bin for inflation to be 2 percent or less. Assume that survey respondents attach equal probabilities of 50 percent to these two bins. Now suppose that the baseline forecast density from the BVAR, represented by Markov Chain Monte Carlo draws from the posterior predictive distribution, has 60 percent of forecast draws that exceed 2 percent and 40 percent of draws of inflation of 2 percent or less. 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 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 Tilting to multiple 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 $\mathbf{X}$ falling into each set of bins, but nothing about the interdependence between $\mathbf{X}$ falling into bins from different histograms. Formally, the SPF prescribes targeted values for the marginal bin probabilities $q(A_i^1)$, $q(A_j^2)$, and $q(A_k^3)$, but not for the joint density $q(A_i^1, A_j^2, A_k^3)$.

In the analytics for multiple histograms, we index the histograms by $n = 1, 2, \dots, N$. Each histogram contains $J(n)$ bins that represent a set of mutually exclusive events. Probabilities of falling into the j th bin of the n th histogram under the proposal and targeted histogram are denoted by $p(A_j^n)$ and $q(A_j^n)$, respectively. For ease of exposition, we present the analytics for $N = 3$ histograms with bin indices i, j , and k for the three histograms, respectively. (The extension to arbitrary N is immediate.) The probabilities pertaining to joint occurrences in bins i, j , and k of the three histograms are:

$$p(A_i^1, A_j^2, A_k^3) \equiv p(\mathbf{X} \in A_i^1 \cap \mathbf{X} \in A_j^2 \cap \mathbf{X} \in A_k^3). \quad (10)$$

To decompose the relative entropy objective, we factorize the densities:

$$\begin{aligned} q(\mathbf{X}) &= q(A_i^1, A_j^2, A_k^3) q(\mathbf{X} | A_i^1, A_j^2, A_k^3) \\ &= q(A_i^1) q(A_j^2) q(A_k^3) m^q(A_i^1, A_j^2, A_k^3) q(\mathbf{X} | A_i^1, A_j^2, A_k^3), \end{aligned} \quad (11)$$

$$\begin{aligned} p(\mathbf{X}) &= p(A_i^1, A_j^2, A_k^3) p(\mathbf{X} | A_i^1, A_j^2, A_k^3) \\ &= p(A_i^1) p(A_j^2) p(A_k^3) m^p(A_i^1, A_j^2, A_k^3) p(\mathbf{X} | A_i^1, A_j^2, A_k^3), \end{aligned} \quad (12)$$

where we have introduced “interaction coefficients” that capture the dependence structure across histograms:

$$m^q(A_i^1, A_j^2, A_k^3) \equiv \frac{q(A_i^1, A_j^2, A_k^3)}{q(A_i^1) q(A_j^2) q(A_k^3)}, \quad (13)$$

and similarly for $m^p(A_i^1, A_j^2, A_k^3)$.

These interaction coefficients are measures of association and capture deviations from independence. Formally, the interaction coefficients are Radon-Nikodym derivatives that describe the change of measure between the product of the marginal distributions and the joint distribution. Throughout, we pose the tilting problem on the positive target support: any SPF bin with $q(A_i^n) = 0$ forces zero tilted mass on every cross-histogram event that involves it. Such events retain zero weight throughout and drop out of the objective, so the interaction coefficients in (13) and the relative-entropy terms that follow are understood to range over cross-histogram events built from bins with $q(A_i^n) > 0$. Equivalently, the multiplier of any zero-probability target bin is held at zero, which our implementation obtains automatically: the augmented proposal density keeps the denominator of each multiplier update strictly positive, so a zero target numerator yields a zero multiplier without any special-casing. From the factorizations in (11) and (12), we obtain a linear decomposition of relative entropy:

$$D(q(\mathbf{X}) \parallel p(\mathbf{X})) = \sum_{n=1}^3 D(q(A^n) \parallel p(A^n)) + \Delta_q(m^q(A^1, A^2, A^3) \parallel m^p(A^1, A^2, A^3))$$

with $\Delta_q(m^q(A^1, A^2, A^3) \parallel m^p(A^1, A^2, A^3)) \triangleq E_q \{ D(q(\mathbf{X}|A^1, A^2, A^3) \parallel p(\mathbf{X}|A^1, A^2, A^3)) \}$, (14)

$$\sum_i \sum_j \sum_k q(A_i^1) q(A_j^2) q(A_k^3) m^q(A_i^1, A_j^2, A_k^3) \log \left(\frac{m^q(A_i^1, A_j^2, A_k^3)}{m^p(A_i^1, A_j^2, A_k^3)} \right), \quad (15)$$

and $E_q \{ D(q(\mathbf{X}|A^1, A^2, A^3) \parallel p(\mathbf{X}|A^1, A^2, A^3)) \} \equiv$

$$\sum_i \sum_j \sum_k q(A_i^1, A_j^2, A_k^3) \cdot D(q(\mathbf{X}|A_i^1, A_j^2, A_k^3) \parallel p(\mathbf{X}|A_i^1, A_j^2, A_k^3)). \quad (16)$$

As in the one-histogram case, the first term in (14) is fixed by the choice of bins and 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 to minimize the relative entropy objective in (14) while satisfying the constraints in (7) is such that the relative entropy between the within-cross-event densities is zero: $q(\mathbf{X}|A_i^1, A_j^2, A_k^3) = p(\mathbf{X}|A_i^1, A_j^2, A_k^3) \forall i, j, k$; and analogously for arbitrary N .*

See Appendix B for the proof of this and subsequent propositions and lemmas.

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 $A_i^1 \cap A_j^2 \cap \dots$. Formally, we require that $p(\mathbf{X} \in A_i^1 \cap A_j^2 \cap \dots) > 0$ for all $i, j, \dots$.*

Proposition 1 leaves the interaction entropy $\Delta_q(\cdot||\cdot)$ in (14) as the only remaining component of relative entropy to be affected by the optimal choice of $q(\mathbf{X})$. Interaction entropy is the residual component of relative entropy attributable to changes in the interaction (or association) structure across histograms; we refer to it loosely as the entropic cost of such changes.¹²

We need to choose the joint bin probabilities $q(A_i^1, A_j^2, A_k^3)$, or equivalently the interaction coefficients $m^q(A_i^1, A_j^2, A_k^3)$, to minimize interaction entropy while satisfying the constraints imposed by the targeted marginal histogram probabilities. These constraints require that the joint density q has the correct marginals, i.e., $\sum_j \sum_k q(A_i^1, A_j^2, A_k^3) = q(A_i^1) \forall i$, and analogously for the other histograms. The first-order conditions for minimizing interaction entropy imply that the interaction coefficients under q are proportional to those under p :

$$m^q(A_i^1, A_j^2, A_k^3) = m^p(A_i^1, A_j^2, A_k^3) \cdot \phi_i^1 \cdot \phi_j^2 \cdot \phi_k^3 \quad \forall i, j, k, \quad (17)$$

where ϕ_i^1 , ϕ_j^2 , and ϕ_k^3 are Lagrange multipliers, each for one of the bin-specific marginal constraints. Appendix C states the Lagrangian of the tilting problem and provides the derivation of these first-order conditions.¹³ The ratio of tilted to untilted probabilities, also known as “tilting weights,” can then conveniently be expressed as a cross-product of the rescaled Lagrange multipliers, which we denote by ψ_i^n .¹⁴

$$w_{i,j,k} \equiv \frac{q(A_i^1, A_j^2, A_k^3)}{p(A_i^1, A_j^2, A_k^3)} = \psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3, \quad \text{with } \psi_i^n \equiv \phi_i^n \cdot \frac{q(A_i^n)}{p(A_i^n)}. \quad (18)$$

In terms of the rescaled Lagrange multipliers, the first-order conditions are:

$$\psi_i^1 \cdot \left(\sum_j \sum_k p(A_i^1, A_j^2, A_k^3) \cdot \psi_j^2 \cdot \psi_k^3 \right) = q(A_i^1), \quad \forall i, \quad (19)$$

$$\psi_j^2 \cdot \left(\sum_i \sum_k p(A_i^1, A_j^2, A_k^3) \cdot \psi_i^1 \cdot \psi_k^3 \right) = q(A_j^2), \quad \forall j, \quad (20)$$

$$\psi_k^3 \cdot \left(\sum_i \sum_j p(A_i^1, A_j^2, A_k^3) \cdot \psi_i^1 \cdot \psi_j^2 \right) = q(A_k^3), \quad \forall k. \quad (21)$$

¹²The notion of “entropic cost” applies only loosely, since Δ_q measures relative entropy only against an unnormalized measure. When the histograms are independent under the baseline, $m^p \equiv 1$, the reference measure in Δ_q becomes the product of the tilted marginals and hence a proper distribution, so that Δ_q reduces to the relative entropy of the tilted joint against that product, as utilized in Lemma 1.

¹³For notational convenience, the appendix parameterizes the dual problem directly in the (logged) rescaled multipliers ψ_i^n defined next, and we refer to ϕ_i^n and ψ_i^n as unscaled and rescaled multipliers, respectively.

¹⁴Henceforth we will be referring to ψ_i^n simply as Lagrange multipliers.

Algorithm 1 (Iterative proportional fitting (for $N = 3$)). For $s = 0$ set initial values $\psi_i^n\{s\} = q(A_i^n)/p(A_i^n)$ for $n = 1, 2, 3$. Starting with $s = 1$, iterate over the following:

1. Given $\psi_j^2\{s-1\} \forall j$ and $\psi_k^3\{s-1\} \forall k$, for each bin i of histogram 1, set $\psi_i^1\{s\}$ to satisfy

$$\psi_i^1\{s\} \cdot \left(\sum_j \sum_k \psi_j^2\{s-1\} \cdot \psi_k^3\{s-1\} \cdot p(A_i^1, A_j^2, A_k^3) \right) = q(A_i^1), \forall i. \quad (22)$$

2. Given $\psi_i^1\{s\} \forall i$ and $\psi_k^3\{s-1\} \forall k$, for each bin j of histogram 2, set $\psi_j^2\{s\}$ to satisfy

$$\psi_j^2\{s\} \cdot \left(\sum_i \sum_k \psi_i^1\{s\} \cdot \psi_k^3\{s-1\} \cdot p(A_i^1, A_j^2, A_k^3) \right) = q(A_j^2), \forall j. \quad (23)$$

3. Given $\psi_i^1\{s\} \forall i$ and $\psi_j^2\{s\} \forall j$, for each bin k of histogram 3, set $\psi_k^3\{s\}$ to satisfy

$$\psi_k^3\{s\} \cdot \left(\sum_i \sum_j \psi_i^1\{s\} \cdot \psi_j^2\{s\} \cdot p(A_i^1, A_j^2, A_k^3) \right) = q(A_k^3), \forall k. \quad (24)$$

Each step enforces the corresponding marginal constraint, and the generalization to arbitrary N is immediate. All constraints are assured to hold when the solution has converged at $\psi_i^1\{s\} = \psi_i^1\{s-1\}$, $\psi_j^2\{s\} = \psi_j^2\{s-1\}$, and $\psi_k^3\{s\} = \psi_k^3\{s-1\} \forall i, j, k$.

Without convenient closed-form solutions available, we obtain the solution by iterating over (19)–(21). This procedure is an instance of “Iterative Proportional Fitting” (IPF), which has a longstanding tradition in the statistics literature (Deming and Stephan, 1940; Ireland and Kullback, 1968; Fienberg, 1970); we discuss the relationship to this literature, and what is specific to our setting, in Section 3.6 below. The first-order conditions and the IPF scheme generalize straightforwardly to arbitrary N .

Convergence of Algorithm 1 rests on two distinct results, and holds under Assumption 3. First, whenever the constraint set is non-empty, the tilting problem (7) is a strictly convex program with a unique solution, as discussed by Robertson, Tallman, and Whiteman (2005), Tallman and West (2022), and West (2024). Second, convergence of Algorithm 1 is a classical result from the IPF literature (Ireland and Kullback, 1968; Fienberg, 1970; Csiszár, 1975). Each step of Algorithm 1 performs an I-projection of the current density onto the set of densities satisfying one histogram’s marginal constraints, and cycling

through these projections converges to the I-projection onto their intersection.^{15 16} As documented in the supplementary online appendix, the algorithm converges quite quickly, often in no more than two dozen steps, each taking less than a second to compute.

The cross-product structure of the optimal tilting weights in (18) is very useful for computational efficiency. 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 n th histogram, the number of tilting weights grows by $\prod_{n=1}^N J(n)$, while the number of Lagrange multipliers grows by $\sum_{n=1}^N J(n)$. For the sake of illustration, consider $N = 3$ histograms, each with $J = 10$ bins; the $10^3 = 1,000$ optimal tilting weights $w_{i,j,k}$ are characterized by cross-products of merely $3 \times 10 = 30$ Lagrange multipliers ψ_i^n . This economy is a direct consequence of the entropic (KL) objective: under a generic objective the optimum need not factor as a cross-product, so the linear-in- N characterization is lost, and approaches that work directly on the joint cell vector instead carry a cost that grows with $\prod_{n=1}^N J(n)$ (see, e.g., Lipsitz and Zhao, 1994).

The cross-product structure 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:*

$$D(q(\mathbf{X}) || p(\mathbf{X})) = \sum_{n=1}^N \sum_{i=1}^{J(n)} q(A_i^n) \cdot \log(\psi_i^n) \quad (25)$$

$$= \sum_n \sum_i q(A_i^n) \cdot \log\left(\frac{q(A_i^n)}{p(A_i^n)}\right) + \sum_n \sum_i q(A_i^n) \cdot \log \phi_i^n, \quad (26)$$

where the first term in (26) 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 as much as possible the interaction structure between different histograms under the proposal density p .

¹⁵Convergence pertains to the weights $w_{i,j,k}$ rather than to the multipliers individually. The first-order conditions (19)–(21) and the cross-product $\psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3$ are invariant to rescaling one histogram's multipliers by a constant $c > 0$ and another's by $1/c$, so the ψ_i^n are determined only up to $N - 1$ scale factors while the weights, and hence the tilted density, are unique. The overall scale is pinned by the targets: summing any one of the enforced marginal conditions over its bins yields $\sum_{i,j,k} p(A_i^1, A_j^2, A_k^3) \cdot \psi_i^1 \psi_j^2 \psi_k^3 = 1$, so the tilted probabilities sum to one after every step of Algorithm 1. The remaining $N - 1$ scale factors are inherited from the initial values; the algorithm is equivariant with respect to them, and they are immaterial for the weights. The same indeterminacy is present in the two-way problem of Ireland and Kullback (1968).

¹⁶The second result presumes positive proposal mass on every cross-histogram event (Assumption 3), which ensures that the constraint set is non-empty — the product of the targeted marginals is feasible — and that the multiplier updates remain well defined at every step. As discussed in Section 3.5, the assumption is routinely violated by the sparse Monte Carlo proposal of our application, which is why we tilt an augmented density p in place of the model's simulated density p^0 , restoring the condition by construction.

3.4 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 $p(\mathbf{X})$ generates draws that are independent across histograms. Relatedly, we also describe a feasible choice of tilting weights that targets the product density, $q^{iid}(A_i^1, A_j^2, A_k^3) = q(A_i^1) \cdot q(A_j^2) \cdot q(A_k^3)$, which would be the optimal solution if the proposal density saw the histograms as independent of one another. Continuing with $N = 3$ histograms indexed by i, j, k , the discussion straightforwardly extends to arbitrary N .

Lemma 1. *[Unit interaction coefficients in case of independence] When the original density, p , is identical to the product density of its marginals, the interaction coefficients all equal one:*

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

Lemma 2. *[Independence tilt] A feasible—though not necessarily entropy-minimizing—choice of tilting weights implements the product density, q^{iid} :*

$$w_{i,j,k}^{iid} \equiv \frac{q^{iid}(A_i^1, A_j^2, A_k^3)}{p(A_i^1, A_j^2, A_k^3)}, \quad \text{with} \quad q^{iid}(A_i^1, A_j^2, A_k^3) \equiv q(A_i^1) \cdot q(A_j^2) \cdot q(A_k^3). \quad (27)$$

It is straightforward to verify that all of the resulting marginal bin probabilities satisfy the targeted (marginal) densities, $q(A_i^n) \forall i, n$.

We refer to this choice of tilting weights described in (27) 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 of 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 $w_{i,j,k}^{iid} = w(A_i^1) \cdot w(A_j^2) \cdot w(A_k^3)$, where $w(A_i^n) \equiv q(A_i^n)/p(A_i^n) \forall i, n$, which implements the product density q^{iid} defined in (27). Equivalently, the first-order conditions then admit the normalization in which all (unscaled) Lagrange multipliers equal one, $\phi_i^1 = \phi_j^2 = \phi_k^3 = 1$. (As the multipliers are identified only up to histogram-specific scale factors, this is a normalization; the tilted probabilities and weights are unique.)*

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} discussed in Lemma 2.

3.5 Computational challenges and implementation

Henceforth, to refer to cross-histogram events compactly, we collect the bin indices of the N histograms in the multi-index $\mathbf{c} \equiv (c_1, c_2, \dots, c_N)$, with $1 \leq c_n \leq J(n)$, so that $\mathbf{c}$ labels the cross-histogram event $A_{c_1}^1 \cap A_{c_2}^2 \cap \dots \cap A_{c_N}^N$. With slight abuse of notation, we write $p(\mathbf{c})$ for the probability that the proposal

assigns to this event, and $p(\mathbf{X}|\mathbf{c})$ for its conditional density within the event (and analogously for q). The set of all cross-histogram events is denoted $\mathcal{C}$, with cardinality $|\mathcal{C}| = \prod_{n=1}^N J(n)$.

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 proposal, $p(\mathbf{c})$, they do not depend on the within-bin densities, $p(\mathbf{X}|\mathbf{c})$. As such, the entropic tilting problem with histogram constraints has a compact structure.

However, in practice, entropic tilting to histograms also faces 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 capture 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.

3.5.1 Sparse proposals and failure to meet the sufficient condition

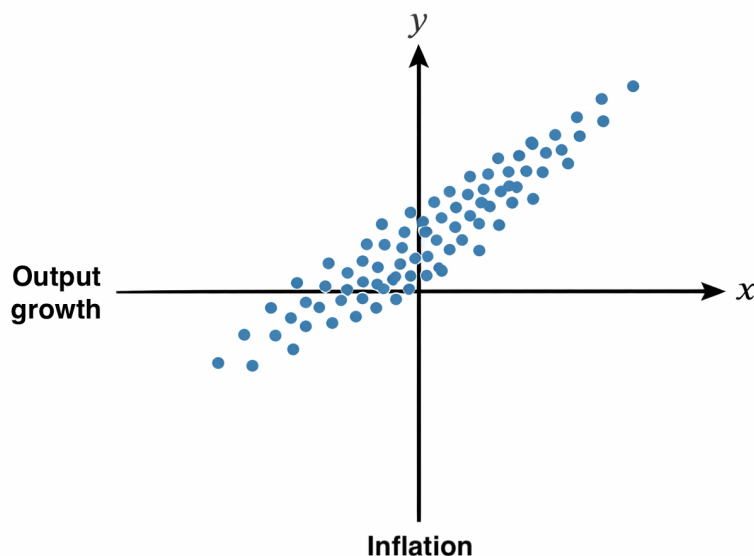
To appreciate the conceptual relevance of the sufficient condition in Assumption 3 we consider the following schematic example. Figure 1 displays a setting with two histograms, one for inflation and one for output growth, that have two bins each. The proposal density provides draws for high and low inflation, as well as high and low output growth; however, it lacks joint occurrences of low inflation and high growth. 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 inflation and high growth. Having zero mass for cross-histogram events is an instance of sparsity in the proposal density p .

Since the SPF puts no explicit target on the combination of low inflation and high 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 problem satisfying all constraints will not be feasible if the sum of the targeted probabilities for low inflation and high growth exceeds 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 can be computationally very appealing.¹⁷ In fact, sparsity becomes a computational necessity as the number of histograms

¹⁷Of 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.

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 into two bins. Blue dots represent draws from the proposal density and there are no draws in the lower right quadrant of low inflation and high growth.

and bins per histogram 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 large number of events $|\mathcal{C}|$ poses significant challenges for processing the (discrete) densities $p(c)$ and $q(c)$. Most notably, the memory requirement of an array of numbers of size 10^{10} is about 80 GB, and simultaneously processing $p(c)$ and $q(c)$ 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 the elements of p are zero, and in practice we find that sparsity can extend to more than 90% of $p(c)$, 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 is 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 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, simulation-based density approximations will fail to deliver draws for some 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 different variables covered by other histograms. Section 3.5.3 below develops this point further, arguing that the observed sparsity is a structural implication of the concentration of the BVAR's predictive density rather than an artifact of an undersized simulation, and that our procedure asks Monte Carlo convergence only of objects the draws can and do deliver. That said, when outcomes are independent and unbounded, the proposal should deliver positive mass for every cross-histogram event 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 of 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 c , as described in the remainder of this section.

3.5.2 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 , and a uniform density for the occurrence of each cross-histogram event c :

$$p(c) = (1 - \omega) p^0(c) + \omega \cdot \frac{1}{|\mathcal{C}|}, \quad (28)$$

where p^0 denotes the original proposal density (say, from the BVAR). The analytics of the preceding sections apply to the augmented density p rather than p^0 .

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 sees the histograms as mutually independent makes cases of large $|\mathcal{C}|$ computationally tractable, since it enables us to lever the insights of Proposition 3 to compute the uniform component's contribution to the probability updates separately for each histogram and independently of one another.

The uniform component in the mixture of (28) leaves the within-joint-event density, $p(\mathbf{X}|c)$, unspecified. We complete it in a way that stays as close as possible to the original proposal: formally, the augmented within-event density is the mixture of $p^0(\mathbf{X}|c)$ and a uniform-within-bin construction, with mixture weights proportional to the two components' contributions to $p(c)$ in (28), as detailed in Appendix E. For cross-histogram events with $p^0(c) > 0$, the weight on the uniform component, $(\omega/|\mathcal{C}|)/p(c)$, is negligible, and the draws generated by p^0 dominate. The augmented proposal is,

¹⁸In our application, we set $\omega = 10^{-4}$. With $|\mathcal{C}| \approx 10^{10}$ that leaves $\omega/|\mathcal{C}|$ just above machine accuracy.

of course, motivated by the occurrence of cases where the original proposal density does not have mass, $p^0(\mathbf{c}) = 0$. For these cases, we generate within-bin values by projections of $\mathbf{X}_t$ onto uniformly drawn values from the bins involved in the cross-histogram event, as detailed in the appendix.

While we find use of the augmented density necessary to obtain an entropic tilting solution 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 typically very small except for the onset of COVID-19. As shown in Figure A.6 in the supplementary online appendix, the mass assigned to zero-mass events rarely exceeds 10%, and the large amounts of mass (of over 60%) assigned in early 2020 are driven by the need to account for unemployment rate histograms in a brief, extremely unusual period.¹⁹ In any event, as shown in Section 5, the tilted forecasts — computed with padding where needed — improve on the accuracy of the untilted BVAR.

3.5.3 Sparsity reflects concentration, not under-sampling

The prevalence of cross-histogram events without draws might be seen as suggesting that our simulated proposal densities are merely under-sampled. At the population level, the BVAR’s predictive density is continuous and strictly positive, so no cross-histogram event has probability exactly zero: emptiness of a cell is purely a finite-sample phenomenon. The accurate description of the simulated proposal is therefore not that the underlying density “is sparse” but that it is *concentrated*. Dependence among the targeted outcomes — persistence of the same variable across adjacent forecast years, and comovement of output growth, unemployment, and inflation — packs probability into a small effective number of cross-histogram events. This concentration produces sparse occupancy even with a large, but finite, number of draws. The concentration is a property of the predictive distribution itself, reproducible across simulation runs; it is not an accident of any particular simulation.

This assessment can be made quantitative. Let $H(\cdot)$ denote Shannon entropy, and let $p^{0,iid}(\mathbf{c}) \equiv \prod_{n=1}^N p^0(A_{c_n}^n)$ be the product of the proposal’s histogram marginals — the maximum-entropy distribution over cross-histogram events consistent with those marginals.²⁰ The *total correlation* of the proposal’s cross-histogram distribution is

$$TC(p^0) \equiv D(p^0(\mathbf{c}) || p^{0,iid}(\mathbf{c})) = \sum_{n=1}^N H(p^0(A^n)) - H(p^0(\mathbf{c})) \geq 0. \quad (29)$$

Cross-histogram dependence shrinks the effective number of cross-histogram events carrying probability mass — measured by the exponential of the entropy, $\exp(H(p^0(\mathbf{c})))$, which equals the number of equiprobable cells consistent with that entropy — by the factor $\exp(TC(p^0))$ relative to independent marginals. Occupancy at a finite number of draws is in turn a functional of the cell probabilities: as de-

¹⁹In such episodes, forecast functionals of the tilted density partly reflect the auxiliary within-event construction of Appendix E — including the finite outer-bin convention of the supplementary online appendix — used to populate events without BVAR draws; padding these events is what preserves the exact matching of the targeted marginal bin probabilities.

²⁰The Shannon entropy of $p^0(\mathbf{c})$ is $H(p^0(\mathbf{c})) \equiv -\sum_{\mathbf{c}} p^0(\mathbf{c}) \log p^0(\mathbf{c})$, and for relating the relative entropy between p^0 and $p^{0,iid}$ to the difference between their entropies we can exploit the fact that $p^0(A_{c_n}^n)$ are the marginals of $p^0(\mathbf{c})$.

tailed in Appendix D, the expected number of empty cross-histogram events in a simulation of K draws equals $\sum_c (1 - p^0(c))^K$, a property of the predictive density and the draw count rather than of any particular simulation run. In our application the cell probabilities are, of course, not available in closed form, so this expression cannot be evaluated directly; a verifiable implication is instead that the occupancy pattern should be invariant across independently seeded simulations of different sizes.²¹

More specifically, the distinction between concentration and under-sampling is verifiable, because the two have opposite empirical signatures. If sparsity were an artifact of under-sampling a diffuse density, increasing the number of draws would populate previously empty cross-histogram events and shift the distribution's mass profile. A concentrated density instead produces an occupancy pattern that broadly persists across choices of the draw count. Section II of the supplementary online appendix documents the latter: across simulations that triple the total number of draws (moving along the number of retained MCMC draws, with independent seeds and hence fully disjoint draw sets), the number of occupied cross-histogram events, the share of empty events, and the total correlation (29) essentially coincide, while the targeted marginal bin probabilities and the joint distribution over commonly covered cross-histogram events agree up to sampling noise.

The remaining question is what our procedure asks of the simulated proposal, and here the entropic-tilting construction itself supplies the answer. We do not estimate a $|\mathcal{C}|$ -cell joint distribution from the model draws — with $|\mathcal{C}| \approx 10^{10}$ against fewer than 10^6 draws, that would indeed be hopeless. The tilt instead enforces $\sum_{n=1}^N J(n)$ marginal constraints, characterized by as many multipliers, and it treats the proposal's joint structure in the most conservative way possible. Specifically, by Proposition 1, the within-event conditional densities are inherited unchanged from the BVAR, and the cross-product form of the optimal weights leaves every cross-histogram odds ratio of the proposal exactly intact. Accordingly, the tilt rescales bin margins while preserving the proposal's association structure. Where the BVAR simulation is silent — cross-histogram events without draws — the uniform, cross-histogram-independent component of the augmented density (28) is precisely the maximum-entropy default: no association is imposed beyond what the BVAR draws and the survey constraints supply. In this precise sense, the tilted density is the *minimal* modification of the (augmented) proposal required by the survey constraints: within the simulated support, the BVAR's association structure is preserved exactly, while in events without draws the tilted association reflects only the uniform padding and the constraints themselves. The joint is characterized — as the relative-entropy projection of the augmented proposal onto the constraint set — rather than estimated cell by cell.

Monte Carlo convergence is accordingly owed at the level at which our claims are made: the targeted marginal bin probabilities, each sampled with an expected draw count of $K p(A_i^n)$ — of order $K/J(n)$ for bins carrying non-negligible mass, while low-mass bins are exactly the ones assigned small target probabilities — and the reported forecast functionals (moments, quantiles, and scores), which weight

²¹For the sake of illustration, consider a bivariate normal with outcomes binned into 10×10 cells. In this case, the cell probabilities — and hence the expected share of empty cells at any draw count — are available in closed form, and that share is sizable and increasing in the correlation coefficient. In this small-scale example, sparsity can be computed analytically rather than simulated, demonstrating that sparse occupancy is a direct consequence of dependence and a finite draw count in a setting where under-sampling is ruled out by construction.

cross-histogram events by their probability mass and therefore inherit the precision of the well-sampled, high-mass region. Section II of the supplementary online appendix shows that these objects converge across draw sizes throughout the evaluation sample, including the COVID-19 origins. Moreover, whenever the simulated cross-tabulation is small relative to the draw count — for the single-variable target sets throughout the sample, whose total cell counts fall below K , and even for the set of all current- and next-year histograms in the early part of the sample, when its cross-tabulation comprised only on the order of 10^2 occupied events — the simulated *joint* distribution itself converges across draw sizes up to sampling noise. The obstacle to such a demonstration for the full target set is its cell count, not a property of the tilting procedure. Finally, Section III of the supplementary online appendix verifies that our headline forecast-accuracy results are essentially unchanged when the proposal density is generated from a much shorter MCMC chain (one third of the baseline number of draws).

3.6 Contributions relative to the IPF literature

Having laid out our method, we can now make precise the contributions anticipated in Section 2. As noted there, the iterative scheme in Algorithm 1 is an instance of IPF, and the multiplicative cross-product form of the optimal weights in (18) together with the convergence of the iteration are classical results (Deming and Stephan, 1940; Ireland and Kullback, 1968; Fienberg, 1970; Csiszár, 1975). Our framework extends this classical work in three respects.

First, whereas the classical literature operates on finite tables of cell probabilities estimated from data, our baseline is a continuous multivariate predictive density and our targets are externally supplied survey histograms. Proposition 1 bridges these settings: because the optimal tilt leaves the within-cross-event conditional densities unchanged, the continuous problem reduces to a discrete problem in the joint bin probabilities—precisely the cross-product problem that Ireland and Kullback (1968) solve. The classical IPF machinery thus applies to the bin probabilities, while the within-bin shape of the predictive density is inherited unchanged from the baseline model.

Second, our N histograms target distinct, generally nonlinear transformations of a shared outcome vector at different forecast horizons. The relevant marginal constraints therefore live in different spaces and must be combined through the joint cross-histogram events $A_i^1 \cap A_j^2 \cap A_k^3 \cap \dots$ introduced above. The interaction entropy of Section 3.3, which the optimal tilt minimizes, is the continuous-density analogue of the association structure that classical IPF preserves in a contingency table. What differs is that the table is not primitive data: each event A_i^n collects the outcomes that transformation n maps into a given bin, and the table's cells are the intersections $A_i^1 \cap A_j^2 \cap \dots$ of one such event per histogram; the baseline interaction coefficients m^p are likewise a functional of the model's joint predictive density rather than of observed counts. Reconciling constraints that live in different spaces is, in this sense, an object the classical problem never has to construct.

Third, the classical convergence and uniqueness results require strictly positive cell counts (Ireland and Kullback, 1968). In our Monte Carlo setting, by contrast, the joint cross-histogram space is populated only by a finite set of model draws, so many cross-histogram events that may need positive mass to satisfy the marginal constraints contain no draws. We confront this sparsity directly through the ag-

nostic padding of the baseline density described above and detailed in Appendix E. Extensions of IPF to incomplete tables with structural zeros do exist (Fienberg, 1972), but they address zeros in observed count data rather than the simulation-based sparsity that is our concern.

Finally, the cross-product structure of the optimal weights also makes the problem computationally tractable as the number of histograms grows: as discussed in Section 3.3, the tilt is characterized by $\sum_{n=1}^N J(n)$ Lagrange multipliers rather than the $\prod_{n=1}^N J(n)$ joint bin probabilities, an economy that direct constrained-optimization or reparameterization approaches to marginal-matching (Lipsitz and Zhao, 1994) do not share.

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 quarterly U.S. Survey of Professional Forecasters.²² 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.²³ 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. 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. 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. Time series of the SPF bins for real GDP growth, the unemployment rate, and core PCE inflation are shown in Figure A.2 in the supplementary online appendix. For example, the SPF lowered the histogram ranges for the unemployment rate a few years after the Great Recession ended, widened them sharply with the COVID-19 pandemic, and narrowed the unemployment

²²The Federal Reserve Bank of Philadelphia took over management of the SPF in 1992 and publishes the survey forecasts on its website.

²³Figure A.1 in the supplementary online appendix shows the histogram forecasts for GDP growth published in the SPF for 2025Q3.

ranges a few years later.

Throughout, we collect probabilistic forecasts for the average SPF forecaster (“consensus histogram”), which is consistent with treating the ensemble of individual forecasts as 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).²⁴ 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. Figures A.3 through A.5 in the supplementary online appendix compare 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 taking the lower and upper bounds of each bin as the location of the probability mass within that bin.²⁵ 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 time.

For the 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%.²⁶ (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 the 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.²⁷

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 unemployment rate, core PCE inflation, and the 5-year

²⁴Studies 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.

²⁵For the open-ended lowest and highest bins, whose literal endpoints are $-\infty$ and ∞ , we use the finite outer limits of Appendix I(a) in place of the infinite ones, so that the mean bounds remain finite whenever a tail bin carries positive probability.

²⁶For a median falling in an open-ended outer bin, the midpoint uses the assumed outer limits of Appendix I(a).

²⁷Nevertheless, 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).

Treasury yield.²⁸ 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).²⁹ 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 2025Q1. 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 1996Q1, 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:

$$y_t = \Pi_0 + \Pi(L)y_{t-1} + v_t, \quad v_t \sim N(0, \Sigma_t), \quad (30)$$

where y_t is a vector of N observables, $\Pi(L) = \sum_{i=1}^p \Pi_i L^{i-1}$ is a p th-order lag polynomial of VAR coefficients, and v_t denotes the VAR's residuals. We denote the vector of stacked coefficients contained in $\{\Pi_i\}_{i=0}^p$ as Π . The model is specified with non-conjugate priors for Π and Σ_t .

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:

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

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:

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

²⁸Quarterly 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.

²⁹In 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.

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 .³⁰ The outlier $o_{j,t}$ has a two-part distribution that distinguishes regular observations with $o_{j,t} = 1$ from outliers for which $o_{j,t} \geq 2$.³¹ Outliers in variable $j = 1, \dots, N$, occur with probability π_j and the distribution:

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

where $U(2, 20)$ denotes a uniform distribution with support between 2 and 20.

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 π_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 15,000 retained draws, obtained by sampling a total of 22,500 draws and discarding the first 7,500. For each retained draw of parameters in the Gibbs sampler, we produce 50 draws of forecasts.³² Accordingly, all results in the paper are

³⁰Stock and Watson (2016) first used this outlier specification with stochastic volatility in a multivariate unobserved component model of inflation.

³¹The 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 $o_{j,t}$).

³²This approach builds on the ideas of Waggoner and Zha (1999). At every Gibbs sampler node, we employ 50 draws of

based on a total of $K = 15,000 \times 50 = 750,000$ draws of forecasts from the BVAR's posterior predictive distribution. Section III of the supplementary online appendix documents that our results are robust to generating the proposal density from a much shorter MCMC chain of 5,000 retained parameter draws (250,000 forecast draws in total).

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 and predictive densities tilted to the SPF histograms. As described in Section 4.1, in our baseline case as available we include SPF histograms for GDP growth, unemployment, and inflation, for horizons from the current year through three years ahead (the “All” set of Table 1). 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 Section 3. Of course, a single given histogram (one variable at one horizon) has multiple bins to hit. Histograms for additional variables or horizons add to the bins to hit. The timing and horizons of BVAR forecasts are aligned with SPF survey rounds.

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.

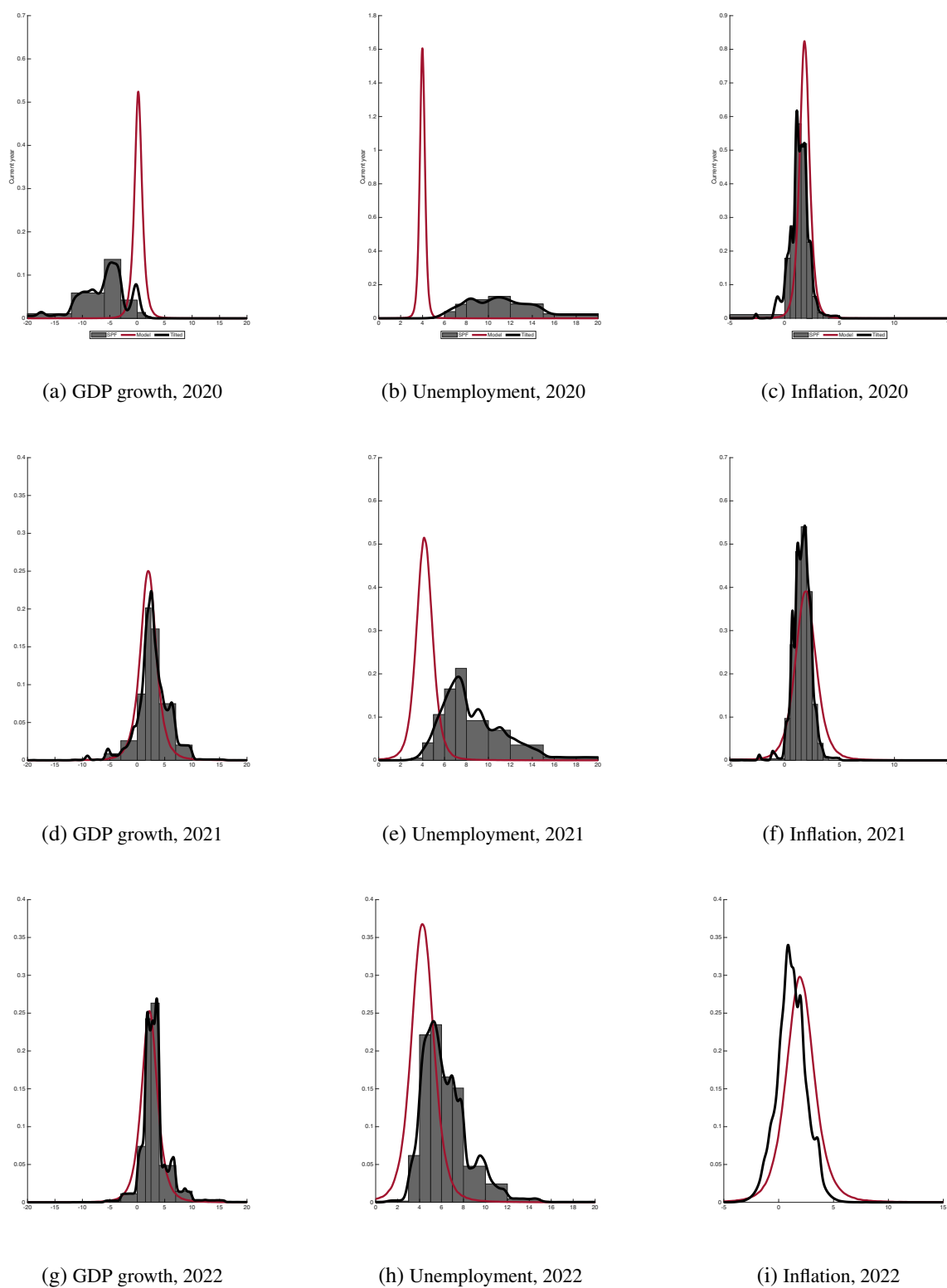
5.1 Visual fit to SPF histograms

We begin by illustrating the differences between raw model output and SPF histogram forecasts for a forecast origin of 2020Q2. (The supplementary online appendix provides results for two additional forecast origins, 2009Q1 and 2021Q2.) Figures 2 and 3 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.

In this example, reflecting the unprecedented circumstances associated with the outbreak of the COVID-19 pandemic, the SPF forecasts differ sharply from the BVAR forecasts, such that entropic

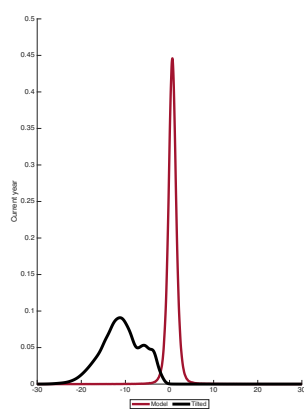
the volatility components and VAR shocks over the forecast horizons, to better balance computational costs against the degree of accuracy in the Monte Carlo approximation of the VAR's predictive likelihood.

Figure 2: Targeted variables, tilted and untilted predictive densities in 2020Q2

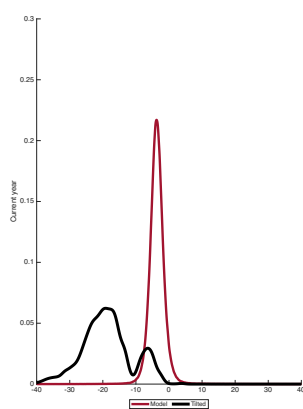


Note: Model-based densities in red, tilted densities and SPF histograms in black.

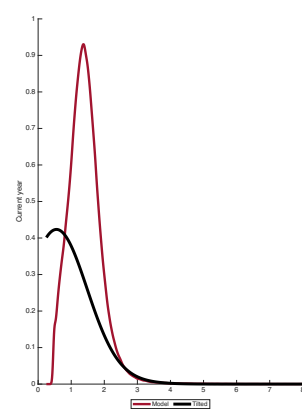
Figure 3: Non-targeted variables, tilted and untilted predictive densities in 2020Q2



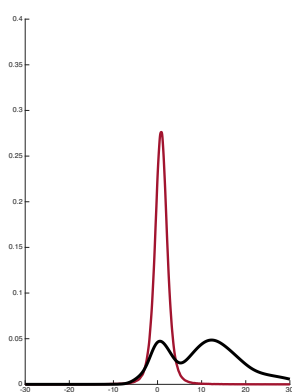
(a) Payroll employment, 2020



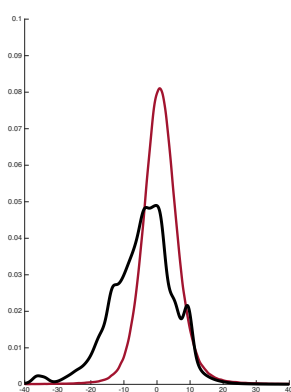
(b) Investment, 2020



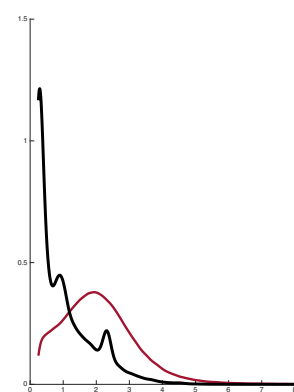
(c) 5-year yield, 2020



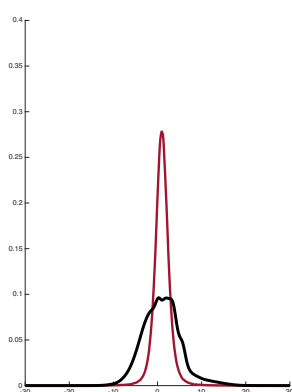
(d) Payroll employment, 2021



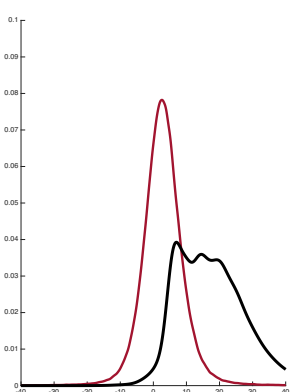
(e) Investment, 2021



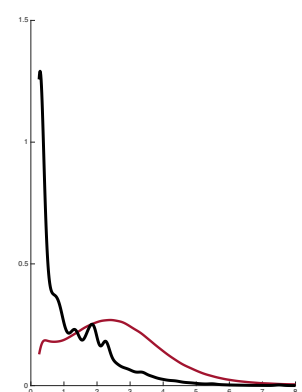
(f) 5-year yield, 2021



(g) Payroll employment, 2022



(h) Investment, 2022



(i) 5-year yield, 2022

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

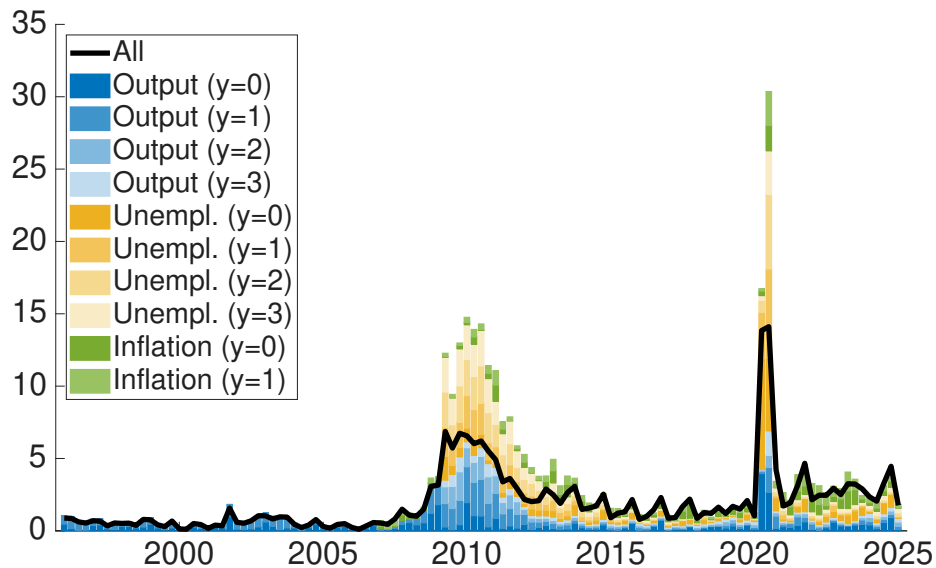
tilting to the SPF histograms has a significant impact on the model's forecasts. More specifically, starting with basic features of the SPF histograms and the BVAR's untilted forecasts, the baseline model, the BVAR-SVO-t specification, produces smooth predictive densities. In contrast, the SPF histograms in some instances exhibit mass concentrated in specific bins, reflecting survey participants' discrete expectations and more timely (higher frequency) information. Most starkly, in 2020Q2, following COVID-19's outbreak, the SPF predictive density for GDP growth in 2020 was sharply shifted and skewed to the left, with probability (at the mode) of 40.91 percent on the bin for growth between -6.0 and -3.1 percent and 35.19 percent on the bin for growth between -12.0 and -6.1 percent (panel 2a). While the BVAR's predictive density shows a significant chance of a decline in GDP for the year, the model had a (relatively) 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 2020 to the left. In the case of the GDP forecast density for 2020, entropically tilting the BVAR's forecasts to match the SPF histograms results in predictive densities for GDP growth with two peaks well defined. 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 to minimize the distance between the tilted and untilted BVAR densities.

Tilting to the SPF histograms also impacts the model's forecasts for unemployment and core PCE inflation, as well as the forecasts for employment, investment, and the 5-year Treasury yield. Matching the SPF histograms also sharply raises the BVAR-based unemployment rate forecast, shifting the predictive density to the right and substantially widening it (panel 2b), consistent with the tilt's reduction of the GDP growth forecast. In addition, tilting to the GDP and unemployment forecasts of the SPF significantly reduces the forecasts of growth in employment (panel 3a) and investment (panel 3b). Incorporating the available SPF histogram forecast of core inflation — which put the probability of inflation below 2 percent at about 83 percent — shifts the BVAR's predictive density for inflation in 2020 to the left and widens its dispersion (panel 2c). The tilting of the BVAR's forecasts to the available SPF histograms also leads to a sizable leftward shift in the forecast of the 5-year Treasury yield (panel 3c).

Forecasts for the year-ahead horizon, the year of 2021, also show significant impacts of entropically tilting the BVAR's forecasts to the SPF histograms, with some of the impacts on 2020 forecasts persisting. In the case of GDP growth, with the SPF histogram reflecting participants' view that growth would snap back from the pandemic-caused drop, tilting causes the BVAR's predictive density to shift somewhat to the right. But the SPF histograms called for unemployment to remain elevated, such that the entropic tilting sharply moves the BVAR's predictive density for unemployment in 2021 to the right (this density provides another example of multiple peaks). In 2021, as in 2020, tilting to the SPF forecasts somewhat shifts to the left the BVAR's predictive density for inflation. While the tilting widens the model-based forecast dispersion for inflation in 2020, it clearly narrows the forecast dispersion in 2021. The BVAR's forecasts for two years ahead continue to show some impact of tilting to SPF histograms (panels 2g, 2h, 2i, 3g, 3h, and 3i).

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.

Figure 4: KLIC decomposition over time



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

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 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 4 provides the time series of total KLIC distance (black line) and contributions (colored bars) from individually tilting to each histogram. The contributions are defined by the first term of equation (26)’s decomposition of relative entropy, which sums the relative entropies between the marginal distributions for each histogram. Since the BVAR’s proposal does not view SPF histograms as independent, the total KLIC distance is not simply the sum of the individual contributions. The difference between the total KLIC and the sum of the individual contributions reflects the (optimized) interaction entropy discussed in Proposition 2. 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- 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

Table 1: Tilting-target sets

Set	Label	Variables and year-ahead horizons			Max. N
		GDP	UNRATE	COREPCE	
All histograms (default)	All	0, 1, 2, 3	0, 1, 2, 3	0, 1	10
Current & next year	y 0+1	0, 1	0, 1	0, 1	6
GDP growth only	GDP	0, 1, 2, 3	—	—	4
Unemployment only	U	—	0, 1, 2, 3	—	4
Core PCE inflation only	Inf	—	—	0, 1	2

Notes: “Label” gives the abbreviation used in tables and figure legends. Horizons are annual, indexed by years ahead (0 = current year); “—” marks a variable not in the set. Each set comprises only the histograms available at a given forecast origin; availability expands over the sample (current- and next-year GDP from 1996Q1, core PCE from 2007Q1, the longer GDP horizons and all unemployment histograms from 2009Q2 — see Section 4.1). “Max. N ” is the set’s size once all its histograms are available; under the “All” set the count grows from 2 (1996Q1–2006Q4) to 4 (2007Q1–2009Q1) to 10 (2009Q2 onward).

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-19’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.

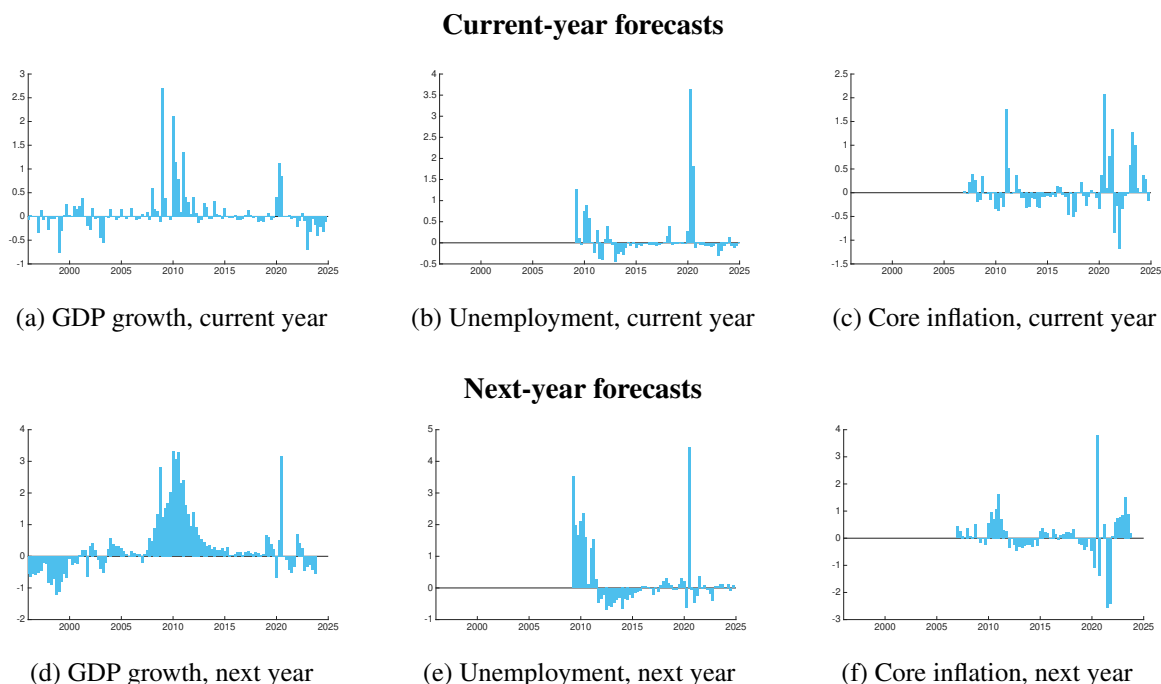
5.2 Forecast accuracy evaluation

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 (the “All” set of Table 1). In the case of GDP growth, we use the current- 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.

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 that outcomes fall into a given SPF-defined bin, as made by the model (without tilting) and by the SPF histogram forecasts. As a summary measure, we compute

Figure 5: DRPS for forecasts from BVAR-SVO-t model and SPF



Note: The figure shows differences in discrete rank probability scores (DRPS) at different forecast origins. The x-axis represents the forecast origin, while the y-axis indicates the difference in raw probability scores between the BVAR-SVO-t model and the SPF histogram forecasts (BVAR score less SPF score). Positive values indicate better density-forecast performance for the SPF (at a given forecast origin), while negative values indicate better performance for the BVAR-SVO-t model.

discrete rank probability scores (DRPS) for both the BVAR-SVO-t model and the SPF histogram forecasts.³³ The DRPS is defined as $DRPS_t = \sum_{j=1}^J \left(P_t^j - D_t^j \right)^2$, with J indicating the number of bins, P_t^j denoting the cumulative forecast probability from bins 1 through j , and D_t^j denoting an indicator that equals 1 if the outcome falls in any bin equal to or lower than j and 0 otherwise. The lower the score, the better the forecast.

Figure 5 reports the DRPS for the BVAR less that for the SPF, over the samples over which annual forecasts from the SPF are available at a given horizon; the plotted object is thus the difference in raw DRPS levels, not a ratio. The percentage magnitudes reported in the text below are a separate normalization, computed as the average score difference relative to the average score of the model-based forecast over the relevant subsample, consistent with the relative DRPS tables in the supplementary online appendix. For chart readability, we report results for the current year and year-ahead horizons; results for longer horizons are qualitatively similar. Average relative DRPS for all available horizons and various subsamples, with statistical significance assessed by Diebold-Mariano tests, are tabulated in the supplementary online appendix. 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 typically less accurate than the BVAR's annual forecasts (on average by roughly 20

³³Past studies of survey forecasts using the DRPS include Boero, Smith, and Wallis (2011), Clements (2018), and Krüger and Pavlova (2024).

to 40 percent in the normalized terms just described). However, for the period since the Great Recession, the SPF's density forecasts are often more accurate than the BVAR's. For real GDP growth, the SPF's advantage is sizable (roughly 30 to 40 percent). In the case of the unemployment rate and core PCE inflation, the SPF's forecasts show the largest gains relative to the BVAR (along with some losses) through about 2016 and mostly smaller score differences for the remainder of the sample (punctuated by pronounced swings around the COVID-19 pandemic), and 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 often 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 $h = 0$ 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 2 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 generally improves the BVAR's quarterly forecast accuracy, both point (RMSE) and density (CRPS). While the sizes of the gains vary over horizons and across variables, 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). The benefits of bringing the SPF histograms to bear are particularly large for GDP growth at $h = 0$, where the current-year SPF forecast lowers the RMSE by close to half and the CRPS by roughly 25 percent. At medium horizons, the GDP gains are more modest, with RMSE reductions of roughly 15 percent and CRPS reductions of about 20 percent, before fading away at long horizons (with RMSE and CRPS ratios near 1.0 by $h = 15$). Tilting yields larger RMSE benefits (albeit not statistically significant) for forecasting the unemployment rate, lowering the RMSE by roughly 25 to 30 percent at medium horizons (with the CRPS gains a bit smaller, around 15 to 20 percent), and roughly 10 percent at longer horizons before dwindling away by $h = 15$. The gains to forecast accuracy for investment are not as large as those for GDP but broadly comparable, whereas the gains for employment forecasts are larger than those for unemployment at some horizons and smaller at other horizons. Tilting also delivers sizable gains for core PCE inflation at medium horizons, with RMSE reductions of roughly

Table 2: Accuracy of tilted and untilted forecasts from BVAR-SVO-t model: 1996-2025

h	RMSE					CRPS						
	Output	BFI	Unempl.	Payroll	Inflation	5y yield	Output	BFI	Unempl.	Payroll	Inflation	5y yield
0	0.53	0.78	0.36	0.65	0.85	0.97	0.73	0.89	0.56	0.70	0.92	0.98
1	0.72	0.88	0.75	0.67	0.87	0.99	0.76*	0.93	0.75	0.68	0.94	0.98
2	0.79	0.94	0.81	0.66	0.87	0.99	0.76*	0.94	0.83	0.70*	0.95	0.98
3	0.74*	0.91	0.74	0.63	0.83	0.96	0.75*	0.92	0.81	0.67**	0.94	0.96
4	0.87	0.93	0.69	0.76	0.83	0.94**	0.80	0.92	0.79	0.71**	0.93	0.95**
5	0.84	0.90	0.70	0.77**	0.80	0.94**	0.79	0.90	0.80	0.71**	0.90	0.93**
6	0.88	0.93	0.72	0.83**	0.77	0.93***	0.79	0.91	0.81	0.74**	0.88**	0.93**
7	0.86	0.92	0.75	0.86*	0.77	0.93***	0.79	0.88	0.82	0.77**	0.87**	0.93**
8	0.86	0.89	0.79	0.88*	0.77	0.93**	0.80	0.86	0.84	0.79*	0.86**	0.93*
9	0.86*	0.86	0.82	0.90	0.77	0.92**	0.81*	0.83	0.87	0.83	0.87*	0.92**
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	0.88*	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*	0.89	0.97	0.81	0.92**	0.90	0.87*	0.93	0.94	0.89*	0.91*
13	0.96	0.92*	0.91	0.98	0.83	0.92**	0.94	0.90*	0.94	0.97	0.90*	0.91*
14	0.99	0.96	0.94	1.00	0.83	0.93**	0.98	0.95	0.96	1.01	0.90*	0.92*
15	1.01	0.99	0.96	1.01	0.84	0.93**	1.02	0.99	0.98	1.03	0.91**	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.

20 percent and CRPS reductions of about 10 percent, commonly statistically significant for the CRPS. For the 5-year Treasury yield, gains are smaller but highly consistent across horizons, with RMSE and CRPS reductions of about 6 to 7 percent that are statistically significant at most horizons. 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 3 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 of 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 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. For the forecast tilted to all available histograms, the resulting time series of the difference in the cumulative average scores relative to the BVAR's forecast are included in Figure 6 below (which also covers the alternative sets of tilting targets examined later), 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

Table 3: Accuracy of tilted and untilted forecasts from BVAR-SVO-t model: 2007-2025

h	RMSE					CRPS						
	Output	BFI	Unempl.	Payroll	Inflation	5y yield	Output	BFI	Unempl.	Payroll	Inflation	5y yield
0	0.49	0.73	0.35	0.65	0.81	0.98	0.64	0.84	0.49	0.67	0.89	0.99
1	0.69	0.86	0.74	0.67	0.83	1.00	0.67**	0.89	0.71	0.65	0.91	0.99
2	0.76	0.92	0.81	0.66	0.83	0.99	0.65**	0.88	0.80	0.67*	0.92	0.98
3	0.69**	0.87	0.73	0.63	0.78	0.96	0.62**	0.83*	0.77	0.65**	0.90	0.95
4	0.82	0.87	0.68	0.76	0.79	0.93**	0.69**	0.82*	0.74	0.68**	0.89	0.93**
5	0.79*	0.82**	0.68	0.77**	0.75	0.92**	0.70*	0.78*	0.75	0.68**	0.85*	0.91**
6	0.85	0.86	0.70	0.83**	0.72	0.91***	0.71*	0.80	0.76	0.72**	0.82**	0.90**
7	0.84	0.86	0.74	0.86*	0.72	0.91***	0.72*	0.78	0.77	0.73*	0.80***	0.90**
8	0.84	0.83	0.77	0.88*	0.72	0.91***	0.74*	0.77	0.80	0.75*	0.80**	0.90**
9	0.83*	0.78*	0.80	0.89	0.72	0.90***	0.74*	0.73*	0.83	0.78	0.81**	0.89**
10	0.87	0.74**	0.83	0.92	0.73	0.90**	0.77*	0.73*	0.85	0.82	0.81**	0.88*
11	0.88*	0.71**	0.84	0.93	0.74	0.90**	0.78*	0.73**	0.88	0.85	0.82*	0.88*
12	0.90*	0.68**	0.87	0.94	0.76	0.89**	0.82**	0.74**	0.90	0.88	0.84*	0.87**
13	0.93**	0.73**	0.88	0.96*	0.78	0.90**	0.88*	0.79**	0.92	0.92	0.85*	0.88**
14	0.97*	0.84**	0.91	0.98	0.77	0.90**	0.95	0.88**	0.94	0.97	0.85**	0.88**
15	1.00	0.91	0.94	1.00	0.78	0.90**	1.01	0.95	0.97	1.01	0.86**	0.88**

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.

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 softening the model's too-strong forecasts of GDP growth, boosting the model's unemployment rate projections, and edging up the model's relatively low forecasts of core inflation. With the pandemic, the relative score results show some additional gains to tilting to the SPF.

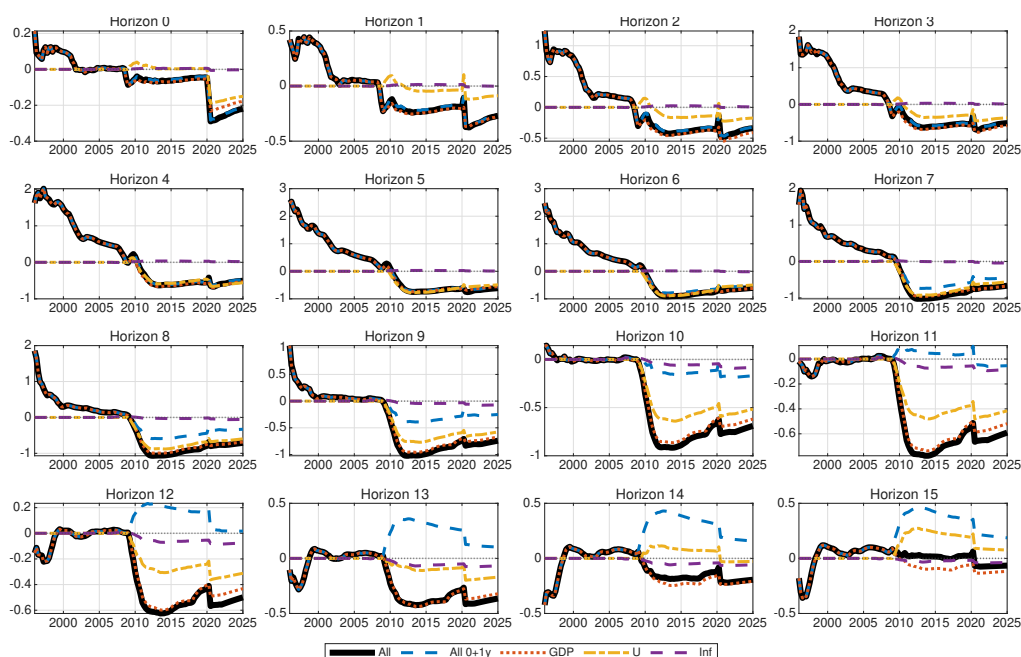
5.5 Tilting to different sets of SPF histograms

In our baseline results, we have focused on tilting BVAR forecasts to a relatively comprehensive set of SPF histograms, with the use of all available histograms having 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 informing BVAR forecasts with survey-based density forecasts in the form of the SPF histograms. However, one could instead choose to tilt to subsets of horizons or variables on the basis of subjective preference or perhaps performance. Many permutations are possible. Rather than engage in an exhaustive search and set of results, we consider a basic range of alternative target sets, each defined in Table 1: (i) the current- and next-year histograms for all three variables ("y 0+1"); (ii) all available GDP growth histograms ("GDP"); (iii) all available unemployment histograms ("U"); and (iv) all available core PCE inflation histograms ("Inf"). When we tilt to a single variable's histograms, we still tilt the BVAR's forecasts for all variables to those histograms.

Starting with the ES measure of joint density accuracy, Figure 6 provides time series of the difference in the cumulative average scores for tilted versus untilted forecasts under the default set of histogram targets and the four alternative sets (Table 1). The supplementary online appendix complements the figure with average relative energy scores for these comparisons, including tests of statistical significance, for both the full sample and 2007-2025 (Tables A.2 and A.3). The approach of tilting to just current- and next-year histograms (blue dotted lines) generally rivals the default approach's accuracy at shorter horizons (from $h = 0$ to 6) but falls short at longer horizons; at horizons of 12 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) nearly matches the success of tilting to all histograms, except that at some horizons it falls short of the default 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 default 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 default 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 A.1 in the supplementary online appendix provides RMSE and CRPS results for the alternative choices of tilting targets, for the full sample of 1996-2025 (results for 2007-2025 are qualitatively similar). All tilting forecasts are compared against the raw BVAR forecasts, and to facilitate comparisons, results for the default tilting set of all variables are provided in the first column of results. Consistent with the ES results, these RMSE and CRPS results indicate that

Figure 6: Energy scores when tilting to different targets over time



Note: Differences in cumulative average energy scores between tilted and untilted densities generated from BVAR-SVO-t and using the alternative target sets of Table 1. Negative values indicate better scores of tilted densities. Quarterly forecast horizons.

some alternative choices of tilting targets can achieve accuracy gains in line with those with the default of using all available histograms, but no other choice matches the default 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 default for most horizons and variables but not quite as accurate for some variables at longer horizons. Tilting to the histograms for a single variable often matches the accuracy of the default tilting set for that variable but without improving the accuracy of other variables as much. For example, tilting to just GDP histograms improves unemployment rate forecasts but by less than tilting to all histograms. Likewise, tilting to inflation histograms alone does not much improve forecasts of the unemployment rate or employment. Altogether, these results indicate that while tilting to fewer histograms can selectively improve on the accuracy of BVAR forecasts, there is no uniform advantage in forecast accuracy over tilting to all available histograms.

5.6 Tilting to moments of densities fitted to SPF histograms

As noted above, some studies have sought to improve model-based forecasts by bringing survey histogram information to bear on BVAR forecasts. These studies differ from ours not only by looking at forecasts for other economies than the U.S., but also by applying entropic tilting only to selected moments fitted onto survey histograms. For example, Galvão, Garratt, and Mitchell (2021) — for the UK — and Banbura, et al. (2024) and Ganics and Odendahl (2021) — for the euro area — have fitted continuous

distributions to summarize histogram information in the form of first or second moments to potentially improve model-based forecasts.

Accordingly, we compare results under this alternative approach to our direct tilting to histograms. The set of variables and forecast horizons in the tilting is the same as in our default tilting to complete SPF histograms (the “All” set of Table 1). At each forecast origin, for each available histogram, we use nonlinear least squares to fit a generalized beta distribution to the cumulative histogram. We compute the mean, variance, and skewness of the fitted beta distribution.³⁴ We then entropically tilt the BVAR forecast distribution to hit those fitted moments, using the conventional tilting methodology of Robertson, Tallman, and Whiteman (2005). We consider the case of tilting to: just means (denoted M); means and variances (MV); and means, variances, and skewness (MVS).³⁵ These variations shed light on the marginal value of the fitted moments of the SPF histograms for adjusting the centering of the BVAR densities (tilting to M), the variance (MV), or additional information on the shape of the density (MVS).³⁶

As a joint measure of forecast accuracy, Table 4 reports average relative energy scores over the 1996-2025 and 2007-2025 samples, comparing (i) forecasts tilted to full histograms, (ii) forecasts tilted to moments from distributions fitted to histograms, and (iii) untilted BVAR forecasts. Two sets of panels in the table compare (i) and (ii) to (iii), and two other sets compare (i) to (ii). The score ratios for fitted moment-tilting compared to the untilted BVAR indicate that the moment information extracted from the SPF histograms consistently improves forecast accuracy, once again with larger gains for 2007-2025 than 1996-2025. Tilting to just fitted means lowers the ES by roughly 5-7 percent at medium horizons (a little more at shorter horizons and less at longer horizons) for these two samples. Compared to tilting to just means, tilting to both fitted means and fitted variances yields larger gains to ES accuracy, with gains against the untilted BVAR (at medium horizons) of roughly 15 percent and 23 percent in the 1996-2025 and 2007-2025 samples, respectively. Our finding that fitted histogram variances can improve BVAR accuracy stands at odds with the findings of Galvão, Garratt, and Mitchell (2021) and Banbura, et al. (2024), but our analysis focuses on the U.S. instead of the UK or the euro area as in these studies.³⁷ Tilting to skewness in addition to means and variances also yields consistent gains on the BVAR’s baseline accuracy, but by less than achieved by tilting to just means and variances.

Now consider the head-to-head forecast accuracy comparison of directly tilting to histograms to tilting to moments of distributions fitted to the histograms (scores that are in the denominator, so that table entries less than 1 mean our approach is more accurate). If the fitted moments used are just means,

³⁴Our implementation follows previous studies such as Galvão, Garratt, and Mitchell (2021).

³⁵From 2009Q2 through the end of our sample, reflecting the SPF’s significant expansion of its set of reported histograms in variables and horizons, we are tilting to significantly more fitted moments than considered in Banbura, et al. (2024), who focused on moments from histograms for two variables and two horizons.

³⁶In this spirit, sticking with our analytical solution to histogram tilting, an alternative option would be to tilt the BVAR forecast to the medians implied by the histograms and 70 percent bands obtained from the 15th and 85th percentiles as measures of uncertainty. In these comparisons, tilting to the median and 15th and 85th percentiles rivals but falls a little short of full histogram tilting, whereas tilting to just the median yields some gains against the raw BVAR but falls well short of full histogram tilting.

³⁷In a different setting, focused on entropic tilting to SPF nowcasts for the U.S., Krüger, Clark, and Ravazzolo (2017) also found some benefit to tilting BVAR forecasts to SPF variances (in their case, measured with past nowcast errors). In this respect, our finding is not without precedent.

Table 4: Relative energy scores: tilting full histograms vs. fitted moments (1996-2025, 2007-2025)

h	full sample										since GFC (2007)					
	... vs. untitled BVAR					Hstgrm. vs. ...					... vs. untitled BVAR			Hstgrm. vs. ...		
	Hstgrm.	M	MV	MVS	M	MV	MVS	Hstgrm.	M	MV	MVS	Hstgrm.	M	MV	MVS	MV
0	0.85	0.93	0.88	0.93*	0.91	0.96	0.92	0.79	0.91	0.83	0.90	0.87	0.87	0.94	0.87	0.87
1	0.89	0.94	0.86	0.91**	0.95	1.03	0.97	0.84	0.91	0.80	0.88**	0.92	0.92	1.04	0.95	0.95
2	0.90	0.94	0.87	0.90**	0.95	1.03	0.99	0.84	0.92	0.80*	0.86***	0.92	0.92	1.05	0.98	0.98
3	0.88	0.94	0.87	0.89**	0.94	1.02	0.99	0.80**	0.91*	0.77**	0.83***	0.88*	0.88*	1.04	0.96	0.96
4	0.89	0.94*	0.89	0.90	0.95	1.00	0.99	0.80**	0.91**	0.79**	0.82***	0.88*	0.88*	1.01	0.97	0.97
5	0.87	0.95	0.88	0.89	0.92	0.99	0.98	0.77**	0.92**	0.79**	0.81**	0.84**	0.84**	0.99	0.96	0.96
6	0.88	0.95	0.87	0.88	0.93	1.01	0.99	0.79*	0.91**	0.77**	0.81**	0.87	0.87	1.02	0.98	0.98
7	0.87	0.95	0.85	0.88	0.91	1.02	0.98	0.79*	0.93*	0.77**	0.81*	0.85*	0.85*	1.03	0.97	0.97
8	0.86	0.96	0.85	0.88	0.89	1.01	0.97	0.78*	0.94	0.77*	0.82*	0.84*	0.84*	1.02	0.96	0.96
9	0.85	0.97	0.85	0.89	0.88*	1.00	0.96*	0.77*	0.95	0.77*	0.83*	0.82**	0.82**	1.00	0.93**	0.93**
10	0.86*	0.97	0.86*	0.90	0.89*	1.00	0.96*	0.78**	0.95*	0.77*	0.84	0.82**	0.82**	1.01	0.92**	0.92**
11	0.88*	0.98	0.88*	0.92	0.89*	1.00	0.96*	0.79**	0.97	0.79*	0.86	0.82**	0.82**	1.00	0.92**	0.92**
12	0.90*	1.00	0.90*	0.93	0.90**	1.00	0.96*	0.81**	0.98	0.81**	0.88*	0.82**	0.82**	1.00	0.92**	0.92**
13	0.92*	1.01	0.92	0.96	0.92**	1.00	0.96*	0.85**	1.00	0.85**	0.92	0.85**	0.85**	0.99	0.92**	0.92**
14	0.96	1.02**	0.96	0.98	0.94**	1.00	0.98	0.91**	1.02	0.90**	0.95**	0.89**	0.89**	1.00	0.95*	0.95*
15	0.99	1.03***	0.98	1.00	0.96*	1.00	0.99	0.96	1.04**	0.95	0.98	0.92**	0.92**	1.01	0.97	0.97

Note: h quarters ahead forecasts. Relative Energy score of each tilting method (joint across all variables), for the evaluation samples indicated in the column headers. Within each sample, the first 4 columns report ratios against the untitled BVAR (in the denominator); the remaining 3 columns report the ratio of the “Hstgrm.” tilt (in the numerator) to each alternative tilt (in the denominator). In case of tilting to fitted moments, the letters “M,” “V,” and “S,” denote tilting to means, variances and skewness, respectively. Evaluation window from 1996Q1 through 2025Q1 (and as far as realized values and tilting targets are available). (full sample: full sample). Evaluation window from 2007Q1 through 2025Q1 (and as far as realized values and tilting targets are available). (since GFC (2007)). 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.

our approach is superior, lowering relative energy scores (at medium horizons) by about 10 (15) percent in the 1996-2025 (2007-2025) sample. This is to be expected given the just-discussed finding that the variance information reflected in the histograms has useful predictive content. Comparing our approach to the alternative of tilting to means and variances of distributions fitted to the histograms, these tilted forecasts yield similar ES accuracy over both samples, with score ratios close to 1 at all horizons and very little evidence of statistically significant differences. With skewness added to the set of fitted moments, tilting to fitted moments yields forecasts slightly to modestly less accurate than those obtained by directly tilting to histograms. Time series of the corresponding cumulative average ES differences, reported in Figure A.11 of the supplementary online appendix, show that the similarity of our direct histograms approach to the fitted means and variances approach applies not only on average over the long samples but also over time. Additionally tilting to fitted skewness can (at some horizons) perform comparably (in the recursive mean differences) up through about 2010 but falls short of direct histogram tilting and mean-variance tilting at longer horizons in the second half of the sample.

Tables A.4 and A.5 in the supplementary online appendix provide RMSE and CRPS accuracy comparisons for each variable over the 1996-2025 and 2007-2025 samples, respectively. To streamline comparisons, these tables compare our approach of direct histogram tilting to fitted moment tilting (in the denominator), and they cover a subset of horizons. Consistent with the ES evidence, the CRPS results show that our approach consistently improves accuracy compared to tilting to just fitted means. But for point forecasts, full histogram tilting yields smaller and less consistent gains over tilting to fitted means. When fitted variances are also used in the tilting, the RMSE and CRPS results show our direct histogram approach to be broadly comparable, with small or modest gains for some variable-horizon combinations and small or modest losses for other combinations. Once again, tilting to fitted skewness does not fare as well: In these RMSE and CRPS results, for most variable and horizon cases, our direct tilting to SPF histograms slightly to modestly improves on forecasts tilted to mean, variance, and skewness moments from distributions fitted to SPF histograms.

Overall, this subsection has shown that our proposed approach to tilting directly to SPF histograms broadly matches but does not exceed the forecast accuracy gains that can be achieved with the extant approach of tilting to first and second moments from distributions fitted to the SPF histograms. At some level, this finding is not surprising: Both approaches employ information on the centering and spread of SPF density forecasts, with our approach taking it directly from the histograms and the fitted moments approach instead extracting information from distributions imposed on the histograms and estimated.

In this context, we would recommend the use of our direct histogram approach for a range of other reasons. Among them, as a conceptual matter noted earlier, by using the histograms directly, one does not impose some external information or judgment on the SPF responses (e.g., the distribution of mass within bins, the parametric shape of a distribution, etc.).³⁸ Moreover, using the histograms directly avoids the need for a step to fit densities to all the SPF histograms. Perhaps more importantly from

³⁸This absence of judgment pertains to the survey side of the problem; as discussed in Section 3.5.2, our implementation does employ auxiliary conventions on the model side to populate cross-histogram events without BVAR draws, which become relevant only when such events receive noticeable tilted mass.

a practical perspective, direct histogram tilting is computationally more robust, avoiding spikes in CPU time that can occur with tilting to fitted moments. As indicated in Figure A.8 in the supplementary online appendix, when tilting to the full set of histograms, direct tilting and tilting to fitted means and variances (MV above) are often comparable in speed; but when MV tilting is slower, it is substantially slower, with execution times rising to a minute or more, precisely around those periods when the KLIC distance indicates histogram information is most informative. For any of the smaller targeting sets, MV tilting is generally slower than direct histogram tilting, by one to two orders of magnitude. If the fitted skewness moment is also used (MVS), the time requirement is larger again, by more than an order of magnitude relative to direct histogram tilting. In addition, at some forecast origins, the entropic tilting to moments from distributions fitted to the histograms can fall short of fully matching the fitted moments (an issue also noted in studies such as Banbura, et al. (2024)).

Finally, even though it is evidently not a significant factor in the accuracy of tilted forecasts, at times the fitted moments-based approach leaves out modest amounts of information in the histograms (also most sharply for nowcasts).³⁹ As discussed in the supplementary online appendix, it is possible to compute the total-variation (TV) distance between the raw SPF bin probabilities (which our direct approach to tilting matches exactly) and the corresponding probabilities of histograms obtained from the fitted moments-tilted predictive density.⁴⁰ In the case of tilting to means and variances, over forecast origins and variables, the median estimate of the TV distance is 0.09, meaning the fitted moment tilt typically misplaces about 9 percentage points of bin probability — representing a discernible but modest mismatch between the moment-tilted histograms and the raw SPF histograms. At times, though, this distance metric is notably higher, corresponding to more visible or even sizable discrepancies. As documented in Section I(d) of the supplementary online appendix, these larger gaps do not mainly reflect a failure of either step of the fitted-moments procedure: the fitted generalized beta reproduces the SPF histogram to within a few percentage points, and, outside the pandemic period, the tilt matches its moment targets essentially exactly. They reflect instead the limits of what a small number of moments can convey about a histogram. The point is sharpest for current-year histograms, which are the easiest to fit parametrically but the hardest to reproduce after moment tilting: As a forecast year progresses and its outcome becomes increasingly known, the SPF histogram concentrates on a few bins, and a density matched only to a mean and a variance cannot recover that near-degenerate shape. In addition, the TV distance is roughly twice as large on average if the fitted moments consist of just the mean or add skewness. These two cases differ in origin, however: Tilting to means alone simply discards the uncertainty information contained in the histogram, whereas adding skewness renders the tilting problem numerically fragile — as shown in the supplementary online appendix, the MVS tilt misses its first-moment targets at many origins throughout the sample, with far larger failures for the higher moments around 2020.

³⁹Together, this finding of omitted information and the occasional difficulty with moment-tilting failing to meet the targeted moments could be seen as reason to prefer our direct histogram tilting (with the padding through a mixture detailed above for dealing with concentration in the BVAR density relative to the many histograms case). In addition, the fact that our proposed approach and tilting to fitted means and variances yield comparable accuracy of quarterly forecasts can be seen as a reassurance that our use of padded draws to solve the tilting problem does not distort forecast accuracy.

⁴⁰This total-variation distance is a complementary measure to the KL distance that can also be calculated.

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 a variety of time series models, including linear VARs, quantile VARs and DSGE models. In addition, our framework for tilting model-based forecasts to histogram forecasts from an external source can readily be applied with other sources of probabilistic forecasts, such as the SPF's recession-probability forecasts. Methodologically, our analytical solution extends the classical iterative proportional fitting procedure for contingency tables (Deming and Stephan, 1940; Ireland and Kullback, 1968; Fienberg, 1970) to forecasting applications in macroeconometrics.

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.

Recent work by Adrian, et al. (2026a) develops the expected misclassification rate (EMR) as a criterion for measuring predictive concordance and for weighting mixtures of scenario forecasts, applied to monetary-policy scenarios in Adrian, et al. (2026b). That criterion governs how competing scenarios are scored against a reference distribution—a different problem from ours, in which a single model-based predictive density is updated to satisfy externally supplied survey constraints. Our analytical results rely directly on the geometry of Kullback–Leibler projections: the closed-form characterization of the single-histogram solution and the cross-product structure exploited by the IPF algorithm are consequences of the exponential-family representation induced by KL minimization. Alternative concordance criteria may be attractive for scenario-weighting problems but need not preserve this analytical and computational structure.

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). Importantly, these gains extend beyond the variables for which SPF histogram targets are available, also improving forecasts for other variables in the system, indicating that survey density information propagates through the joint predictive structure of the multivariate model. 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.

Broadly, our results indicate that survey density forecasts contain system-wide probabilistic information that can be incorporated into multivariate forecasting models in a tractable and nonparametric manner.

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 high GDP growth and low 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 inflation (high or low) and the second index corresponds to GDP growth (low or high), matching the quadrants of Figure 1 read in matrix order — (1, 1) top left, (1, 2) top right, (2, 1) bottom left, (2, 2) bottom right — so that q_{11} is high inflation and low growth, q_{12} is high inflation and high growth, q_{21} is low inflation and low growth, and q_{22} is low inflation and high growth. The marginal probabilities are denoted as $q_{1\cdot}^{\text{INF}} = q_{11} + q_{12}$ for high inflation, $q_{2\cdot}^{\text{INF}} = q_{21} + q_{22}$ for low inflation, $q_{\cdot 1}^{\text{GDP}} = q_{11} + q_{21}$ for low GDP growth, and $q_{\cdot 2}^{\text{GDP}} = q_{12} + q_{22}$ for high GDP growth.

We take it as given that the SPF probabilities sum to one, so that $q_{1\cdot}^{\text{INF}} + q_{2\cdot}^{\text{INF}} = 1$ and $q_{\cdot 1}^{\text{GDP}} + q_{\cdot 2}^{\text{GDP}} = 1$, so that the SPF imposes two linear restrictions on the four entries of q . The zero restriction from the proposal, $q_{22} = 0$ (low inflation and high growth), then implies $q_{21} = q_{2\cdot}^{\text{INF}}$ and $q_{12} = q_{\cdot 2}^{\text{GDP}}$. 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: $q_{11} = 1 - q_{12} - q_{21} = 1 - q_{\cdot 2}^{\text{GDP}} - q_{2\cdot}^{\text{INF}} \geq 0 \iff q_{2\cdot}^{\text{INF}} + q_{\cdot 2}^{\text{GDP}} \leq 1$. As a result, a feasible solution exists only if the sum of the SPF's marginal probabilities for low inflation and high growth does not exceed one.

B Proofs of propositions

Proposition 1. *[Optimal tilting leaves within-cross-event densities unchanged] The optimal choice of q to minimize the relative entropy objective in (14) while satisfying the constraints in (7) is such that the relative entropy between the within-cross-event densities is zero: $q(\mathbf{X}|A_i^1, A_j^2, A_k^3) = p(\mathbf{X}|A_i^1, A_j^2, A_k^3) \forall i, j, k$; and analogously for arbitrary N .*

Proof. The result follows from the ability to choose $q(\mathbf{X}|A_i^1, A_j^2, A_k^3)$ without constraints and separately from $q(A_i^1, A_j^2, A_k^3)$. In addition, the non-negativity of relative entropy, $D(q||p) \geq 0$, 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,

$$D\left(q(\mathbf{X}|A_i^1, A_j^2, A_k^3) \parallel p(\mathbf{X}|A_i^1, A_j^2, A_k^3)\right), \quad \text{this translates to}$$

$$q(\mathbf{X}|A_i^1, A_j^2, A_k^3) = p(\mathbf{X}|A_i^1, A_j^2, A_k^3) \forall i, j, k. \quad \square$$

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:*

$$D(q(\mathbf{X}) \parallel p(\mathbf{X})) = \sum_{n=1}^N \sum_{i=1}^{J(n)} q(A_i^n) \cdot \log(\psi_i^n) \quad (25)$$

$$= \sum_n \sum_i q(A_i^n) \cdot \log\left(\frac{q(A_i^n)}{p(A_i^n)}\right) + \sum_n \sum_i q(A_i^n) \cdot \log \phi_i^n, \quad (26)$$

where the first term in (26) 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 as much as possible the interaction structure between different histograms under the proposal density p .

Proof. By Proposition 1, the optimizing solution leaves the within-cross-histogram event densities unchanged. The relative entropy at the optimum then reduces to

$$D(q(\mathbf{X}) \parallel p(\mathbf{X})) = D(q(A_i^1, A_j^2, A_k^3) \parallel p(A_i^1, A_j^2, A_k^3)) \quad (31)$$

$$= \sum_i \sum_j \sum_k q(A_i^1, A_j^2, A_k^3) \log(w_{i,j,k}) \quad (32)$$

$$= \sum_i \sum_j \sum_k q(A_i^1, A_j^2, A_k^3) \log(\psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3), \quad (33)$$

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

Lemma 1. *[Unit interaction coefficients in case of independence] When the original density, p , is identical to the product density of its marginals, the interaction coefficients all equal one:*

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

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 $N = 3$. $\square$

Lemma 2. *[Independence tilt] A feasible—though not necessarily entropy-minimizing—choice of tilting weights implements the product density, q^{iid} :*

$$w_{i,j,k}^{iid} \equiv \frac{q^{iid}(A_i^1, A_j^2, A_k^3)}{p(A_i^1, A_j^2, A_k^3)}, \quad \text{with } q^{iid}(A_i^1, A_j^2, A_k^3) \equiv q(A_i^1) \cdot q(A_j^2) \cdot q(A_k^3). \quad (27)$$

It is straightforward to verify that all of the resulting marginal bin probabilities satisfy the targeted (marginal) densities, $q(A_i^n) \forall i, n$.

As argued next, the weights $w_{i,j,k}^{iid}$, defined in (27), are the optimal choice when the proposal density sees the histograms as independent of 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 $w_{i,j,k}^{iid} = w(A_i^1) \cdot w(A_j^2) \cdot w(A_k^3)$, where $w(A_i^n) \equiv q(A_i^n)/p(A_i^n) \forall i, n$, which implements the product density q^{iid} defined in (27). Equivalently, the first-order conditions then admit the normalization in which all (unscaled) Lagrange multipliers equal one, $\phi_i^1 = \phi_j^2 = \phi_k^3 = 1$. (As the multipliers are identified only up to histogram-specific scale factors, this is a normalization; the tilted probabilities and weights are unique.)*

Proof. When the histograms are independent under the proposal, we have

$$p(A_i^1, A_j^2, A_k^3) = p(A_i^1) \cdot p(A_j^2) \cdot p(A_k^3) \Rightarrow w_{i,j,k}^{iid} = \frac{q(A_i^1)}{p(A_i^1)} \cdot \frac{q(A_j^2)}{p(A_j^2)} \cdot \frac{q(A_k^3)}{p(A_k^3)}, \quad (34)$$

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

$$\Delta_q(m^q(A^1, A^2, A^3) || 1) = D(q(A^1, A^2, A^3) || q(A^1) \cdot q(A^2) \cdot q(A^3)), \quad (35)$$

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

C Tilting problem and its FOC

This appendix states the Lagrangian of the entropic tilting problem (7) and derives the first-order conditions that underlie the solution reported in Section 3.3: the proportionality between the interaction coefficients of q and p , the cross-product form of the tilting weights in (18), and the system (19)–(21) that is solved by Algorithm 1. As in the main text, we present the case of $N = 3$ histograms with bin indices i, j , and k for the three histograms; the extension to arbitrary N is immediate and noted at the end.

Throughout, we maintain Assumption 3, so that $p(A_i^1, A_j^2, A_k^3) > 0$ for all i, j, k , and the tilting weights $w_{i,j,k} \equiv q(A_i^1, A_j^2, A_k^3)/p(A_i^1, A_j^2, A_k^3)$ are well defined. In our implementation, this condition is assured by use of the augmented proposal density described in Section 3.5.

By Proposition 1, the optimal tilted density inherits the proposal's conditional densities within every cross-histogram event, so that the last term of the decomposition in (14) is zero at the optimum. What remains is a finite-dimensional problem in the joint cross-event probabilities $q(A_i^1, A_j^2, A_k^3)$ or, equiva-

lently, in the tilting weights $w_{i,j,k}$:

$$\min_{\{w_{i,j,k}\}} \sum_i \sum_j \sum_k p(A_i^1, A_j^2, A_k^3) w_{i,j,k} \log(w_{i,j,k}) \quad (36)$$

$$\text{s.t.} \quad \sum_j \sum_k p(A_i^1, A_j^2, A_k^3) w_{i,j,k} = q(A_i^1) \quad \forall i, \quad (37)$$

$$\sum_i \sum_k p(A_i^1, A_j^2, A_k^3) w_{i,j,k} = q(A_j^2) \quad \forall j, \quad (38)$$

$$\sum_i \sum_j p(A_i^1, A_j^2, A_k^3) w_{i,j,k} = q(A_k^3) \quad \forall k. \quad (39)$$

A few observations simplify the problem. First, no separate normalization constraint is required: summing the constraints in (37) over i yields $\sum_i q(A_i^1) = 1$, since the SPF's histogram probabilities sum to one (and analogously for histograms $n = 2, 3$). Second, as noted in Section 3.3, bins to which the SPF assigns zero probability, $q(A_i^n) = 0$, force zero tilted mass on every cross-event they are involved in; these events drop out of the objective and their multipliers are held at zero. The derivation below accordingly applies to the bins with positive target probability. Third, non-negativity constraints on the weights can be omitted because they do not bind at the optimum: the derivative of $x \log(x)$ tends to $-\infty$ as $x \rightarrow 0^+$, so that no retained cross-event with positive proposal mass is optimally assigned zero tilted mass. Fourth, the problem is convex: the objective is strictly convex in the weights, and the constraints are linear. Fifth, given that the (augmented) proposal density assigns positive mass to every retained cross-event (Assumption 3), a strictly positive feasible point always exists — the product density $q(A_i^1) \cdot q(A_j^2) \cdot q(A_k^3)$ satisfies every marginal constraint — so that Lagrange multipliers exist, and the first-order conditions below are both necessary and sufficient for the unique global optimum (see also Robertson, Tallman, and Whiteman, 2005).

For notational convenience, we express multipliers to the constraints (37)–(39) in logged form, as $\log(\psi_i^1)$, $\log(\psi_j^2)$, and $\log(\psi_k^3)$, and the Lagrangian of the problem is:

$$\begin{aligned} \mathcal{L}(\{w_{i,j,k}\}, \{\psi_i^1\}, \{\psi_j^2\}, \{\psi_k^3\}) &= \sum_i \sum_j \sum_k p(A_i^1, A_j^2, A_k^3) w_{i,j,k} \log(w_{i,j,k}) \\ &+ \sum_i \log(\psi_i^1) \left(q(A_i^1) - \sum_j \sum_k p(A_i^1, A_j^2, A_k^3) w_{i,j,k} \right) \\ &+ \sum_j \log(\psi_j^2) (\cdot) + \sum_k \log(\psi_k^3) (\cdot), \quad (40) \end{aligned}$$

where the second and third sums of multiplier terms are defined analogously to the first. Differentiating

with respect to $w_{i,j,k}$, and using $p(A_i^1, A_j^2, A_k^3) > 0$, the first-order conditions are:⁴¹

$$\frac{\partial \mathcal{L}}{\partial w_{i,j,k}} = 0 \implies \log(w_{i,j,k}) + 1 = \log(\psi_i^1) + \log(\psi_j^2) + \log(\psi_k^3), \quad \forall i, j, k \quad (41)$$

such that $w_{i,j,k} \propto \psi_i^1 \cdot \psi_j^2 \cdot \psi_k^3$. Absorbing the constant in (41) into the multipliers, the optimal tilting weights take the cross-product form reported in (18).⁴² After substitution into the marginal constraints (37) we obtain:

$$\psi_i^1 \cdot \left(\sum_j \sum_k p(A_i^1, A_j^2, A_k^3) \cdot \psi_j^2 \cdot \psi_k^3 \right) = q(A_i^1) \quad \forall i, \quad (42)$$

for $n = 1$, and analogously for the other two histograms. This is precisely the system (19)–(21) on which Algorithm 1 iterates.

For arbitrary N , with cross-events indexed by the multi-index $\mathbf{c} = (c_1, \dots, c_N)$, the first-order conditions read $\log(w_{\mathbf{c}}) + 1 = \sum_{n=1}^N \log(\psi_{c_n}^n)$; after the same absorption of the constant, the cross-product representation $w_{\mathbf{c}} = \prod_{n=1}^N \psi_{c_n}^n$ obtains, and substitution into the marginal constraints again yields the first-order conditions solved by IPF.

D Expected occupancy of cross-histogram events

Consider K independent draws from the proposal, each falling into exactly one cross-histogram event, with the probability of event $\mathbf{c}$ given by $p(\mathbf{c})$. In our application, $K = G \cdot P$, where G denotes the number of retained MCMC draws of the model parameters and P the number of forecast paths simulated per parameter draw. Let $K_{\mathbf{c}}$ denote the number of draws falling into event $\mathbf{c}$, so that the occupancy counts across all $\mathbf{c} \in \mathcal{C}$ are jointly multinomial with K trials and event probabilities $p(\mathbf{c})$. A given event is empty if and only if every one of the K draws misses it, which occurs with probability

$$\Pr[K_{\mathbf{c}} = 0] = (1 - p(\mathbf{c}))^K, \quad (43)$$

the binomial probability of zero successes in K trials. Collecting the emptiness indicators $Z_{\mathbf{c}} \equiv \mathbf{1}\{K_{\mathbf{c}} = 0\}$, the set of empty events is $\mathcal{C}_0 \equiv \{\mathbf{c} \in \mathcal{C} : K_{\mathbf{c}} = 0\}$, whose cardinality is $|\mathcal{C}_0| = \sum_{\mathbf{c} \in \mathcal{C}} Z_{\mathbf{c}}$; we write $\mathcal{C}_+ \equiv \mathcal{C} \setminus \mathcal{C}_0$ for the complementary set of occupied, or active, events. By linearity of the expectations operator,

$$E[|\mathcal{C}_0|] = \sum_{\mathbf{c} \in \mathcal{C}} \Pr[K_{\mathbf{c}} = 0] = \sum_{\mathbf{c} \in \mathcal{C}} (1 - p(\mathbf{c}))^K. \quad (44)$$

⁴¹As described in Section 3, note that the first-order conditions acquire content only with $N \geq 2$ histograms. For the single-histogram case, $N = 1$, the constraints alone pin down the solution, $p(A_j) w_j = q(A_j) \forall j$.

⁴²In passing, note that the multipliers are identified only up to $N - 1 = 2$ relative scale factors: replacing ψ_i^n by $\kappa_n \psi_i^n$ with $\kappa_1 \kappa_2 \kappa_3 = 1$ leaves the weights implied by (41) unchanged; correspondingly, one constraint per histogram beyond the first is redundant, since each histogram's constraint set separately implies the same normalization $\sum_{i,j,k} q(A_i^1, A_j^2, A_k^3) = 1$.

Notably, no independence across events is required for (44): the indicators Z_c are strongly dependent — the occupancy counts sum to K — but linearity of expectations holds regardless; the dependence would matter for the variance of $|\mathcal{C}_0|$, not for its mean.

Since $\log(1 - p) \leq -p$, each term of (44) satisfies $(1 - p(c))^K \leq \exp(-K \cdot p(c))$, with near-equality whenever $p(c)$ is small — the relevant regime when individual cross-histogram events carry little mass — so that

$$E[|\mathcal{C}_0|] \approx \sum_{c \in \mathcal{C}} \exp(-K \cdot p(c)). \quad (45)$$

The approximation (45) makes the connection between concentration and sparsity transparent: an event is occupied with high probability once $K \cdot p(c) \gg 1$ and empty with high probability once $K \cdot p(c) \ll 1$, with the transition occurring around $p(c) \approx 1/K$. The expected number of occupied events, $E[|\mathcal{C}_+|]$, is thus, to good approximation, the number of events whose probability exceeds roughly $1/K$ — a property of the proposal density and the draw count, not of any particular simulation run. This threshold reading presumes that the aggregate probability mass of events far below the threshold is small; otherwise, the sub-threshold tail contributes approximately K times its total mass to expected occupancy. A steeply decaying profile of event probabilities — as implied by the concentration documented in Section II of the supplementary online appendix — delivers exactly this negligible tail contribution. When the profile of event probabilities decays steeply below this threshold, as it does for a concentrated joint distribution, increasing K recruits almost no additional events, which is the invariance across draw sizes documented in Section II of the supplementary online appendix.⁴³

E Details of the computational implementation

The IPF iteration of Algorithm 1 and the drawing of $\mathbf{X}$ from $q(\mathbf{X})$ are both implemented with the augmented density (28) as proposal. In our application with all histograms, $|\mathcal{C}| \approx 10^{10}$, which rules out forming any array indexed by all cross-events. Our implementation handles the augmented density exactly by exploiting the sparsity of p^0 and the cross-histogram independence of the uniform component. We store p^0 as a list of pairs $(c, \mathcal{I}(c))$, one for each active cross-event $c \in \mathcal{C}_+$ (i.e., each c with $p^0(c) > 0$, in the notation of Appendix D), where $\mathcal{I}(c)$ collects the indices of the BVAR draws that fall into c .

The first-order conditions in (19)–(21) require, for each histogram n and each bin i , summations of the form $T_i^n \equiv \sum_{c: c_n=i} p(c) \prod_{m \neq n} \psi_{c_m}^m$. Under the augmented density (28), these decompose as

$$T_i^n = (1 - \omega) T_i^{0,n} + \frac{\omega}{|\mathcal{C}|} \prod_{m \neq n} S_m, \quad (46)$$

⁴³Two qualifications apply to our setting. First, the $K = G \cdot P$ simulated draws are not independent — forecast paths sharing one of the G parameter draws are dependent, and the parameter draws themselves are serially dependent along the MCMC chain — so that (43) holds exactly for the independent-draws benchmark, with positive dependence across draws acting like a reduction of the effective sample size. Second, the event probabilities $p(c)$ are not available in closed form in our application, so (44) cannot be evaluated directly; its testable implication is the invariance of the occupancy pattern across independently seeded simulations of different sizes.

where $T_i^{0,n} \equiv \sum_{c: c_n=i} p^0(c) \prod_{m \neq n} \psi_{c_m}^m$ and $S_m \equiv \sum_{j=1}^{J(m)} \psi_j^m$. The sparse contribution $T_i^{0,n}$ is computed only over the active cross-events, using grouping operations whose cost scales with $|\mathcal{C}_+|$ rather than with $|\mathcal{C}|$. The uniform contribution requires only the row sums $\{S_m\}$ and is evaluated at cost linear in N : because the uniform component sees the histograms as mutually independent (cf. Proposition 3), the underlying summation over $\mathcal{C}$ factorizes into $\prod_{m \neq n} S_m$, even though it is taken over all of $\mathcal{C}$. The iteration therefore operates exactly on the augmented density without ever forming an array indexed by $\mathcal{C}$.

Once the ψ_i^n have converged, draws from $q(\mathbf{X})$ proceed in two stages. We first sample a cross-event c from $q(c)$, and then sample $\mathbf{X}$ from $p(\mathbf{X}|c)$ within that cross-event, since $q(\mathbf{X}|c) = p(\mathbf{X}|c)$ at the optimum (Proposition 1). To sample c without ever forming a vector indexed by $\mathcal{C}$, we partition the histograms into batches and draw sequentially batch by batch.⁴⁴ For each batch we compute the marginal probability of q over the joint bins of that batch, conditional on the bins already drawn in earlier batches. This conditional marginal inherits the sparse-plus-uniform decomposition of (46): the sparse part is evaluated only over the active cross-events consistent with the path through the prior batches, and the uniform part exploits the same factorization across histograms.

Conditional on the drawn c , the within-event density $p(\mathbf{X}|c)$ is itself a mixture of the baseline BVAR draws and the uniform-within-bin construction, in proportion to their contributions to the augmented cross-event mass $p(c)$, as defined in (28). With probability $(1 - \omega) p^0(c)/p(c)$ we sample $\mathbf{X}$ from the baseline conditional $p^0(\mathbf{X}|c)$, by picking one of the BVAR draws that fall in c uniformly at random. With the complementary probability $(\omega/|\mathcal{C}|)/p(c)$, $\mathbf{X}$ is generated from the uniform-within-bin component of the augmented density, as described next. For an empty cross-event, with $p^0(c) = 0$, the first probability vanishes and $\mathbf{X}$ is always generated from the within-bin construction; for a heavily populated cross-event it is close to one, so the BVAR draws dominate and the within-bin construction contributes only at the margin.

When the drawn cross-event lies in the zero-mass region of p^0 , the within-bin density must be constructed. For each targeted histogram n , we draw a value $\tilde{f}_n$ uniformly within the bin $A_{c_n}^n$; for the open-ended lowest and highest bins, we draw uniformly over the closed bin obtained under the finite outer-limit convention of Appendix I(a) of the supplementary online appendix, so that $\tilde{f}_n$ is always well defined. We then translate $(\tilde{f}_1, \dots, \tilde{f}_N)$ into a draw of $\mathbf{X}$ in two steps. First, for each targeted variable and forecast year we invert f^n under a constant-growth assumption (described below) to obtain quarterly values $\tilde{\mathbf{X}}_{\text{tgt}}$ for the corresponding forecast quarters of $\mathbf{X}$, with $f^n(\tilde{\mathbf{X}}_{\text{tgt}}) = \tilde{f}_n$ by construction. Second, the remaining elements of $\mathbf{X}$ — non-targeted variables and forecast horizons not covered by any histogram — are drawn from their distribution conditional on the histogram features and the targeted-quarter values. We approximate this conditional distribution as Gaussian and estimate its first two moments, at each forecast origin t , from the BVAR's simulated draws of $\mathbf{X}$. The conditional mean

⁴⁴In our implementation, batches contain three histograms each, balancing the cost per batch (proportional to the joint number of bins in the batch) against the number of batches.

is the OLS projection

$$\hat{\mathbf{X}}_{\text{non-tgt}} = \hat{\alpha} + \hat{B} \cdot (f^1(\mathbf{X}), \dots, f^N(\mathbf{X}), \mathbf{X}_{\text{tgt}}), \quad (47)$$

and the conditional covariance $\hat{\Sigma}$ is the sample covariance of the projection residuals. For linear histogram features, such as the annual-average unemployment rate, the feature is an affine combination of the targeted quarterly values, so that the regressor set in (47) is collinear by construction; the projection is computed by least squares via a QR decomposition of the regressor matrix, which accommodates the redundant columns and leaves the fitted values and residuals unaffected. Evaluating the mean at $\tilde{f}$ and $\tilde{\mathbf{X}}_{\text{tgt}}$, we draw the untargeted block from

$$\tilde{\mathbf{X}}_{\text{non-tgt}} \sim \mathcal{N}(\hat{\mathbf{X}}_{\text{non-tgt}}, \hat{\Sigma}). \quad (48)$$

Sampling from (48), rather than fixing the untargeted coordinates at the projection line (47), preserves the BVAR-implied spread and co-movement among the untargeted coordinates. The resulting draw $\tilde{\mathbf{X}}$ combines $\tilde{\mathbf{X}}_{\text{tgt}}$ in its targeted-quarter slots with $\tilde{\mathbf{X}}_{\text{non-tgt}}$ elsewhere; since f^n depends only on the targeted quarters, $f^n(\tilde{\mathbf{X}}) = \tilde{f}_n$ continues to hold exactly. Inverting f^n is straightforward when the transformation is linear, as is the case for the unemployment rate. For nonlinear f^n the inversion must be solved numerically. For GDP growth in particular, $f_t^n(\mathbf{X}_t)$ is the ratio of average annual GDP levels, which leaves the k forecast quarters of the targeted year underdetermined by $\tilde{f}_n$. We close the system by assuming a constant per-quarter gross growth rate $x > 0$ across these quarters, so that x satisfies $x + x^2 + \dots + x^k = X_{\text{gross}}(\tilde{f}_n)$, where $X_{\text{gross}}(\tilde{f}_n)$ is the residual gross level implied by $\tilde{f}_n$, the levels of already-observed quarters of the target year, and the level of the previous calendar year's average GDP (annual-average growth depends on both years' averages). For target years beyond the current one, the previous year's quarters are themselves partly constructed by this procedure; the inversions are therefore carried out in chronological order of target years, with each year's constructed quarterly levels feeding into the next. This polynomial has a unique positive real root, which we obtain by Newton iteration. Core PCE inflation is handled analogously, with the same constant-growth assumption across the forecast quarters of the targeted year; because $\log(1 + f_t^n(\mathbf{X}_t)/100)$ is linear in $\mathbf{X}_t$, the resulting equation in the per-quarter log inflation rate is linear and no root-finding is needed.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude in order to help with references and literature review, as a coding and editing tool, and to generate the illustrative diagram in Figure 1. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Adrian, Tobias, Domenico Giannone, Matteo Luciani, and Mike West (2025), “Scenario synthesis and macroeconomic risk,” Papers 2505.05193, arXiv.org.
- (2026a), “Predictive concordance for parameter optimisation and mixture synthesis,” Papers 2606.14382, arXiv.org.
- (2026b), “Risks and uncertainty in monetary policy,” Papers 2606.16708, arXiv.org.
- 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.
- 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), “An iterative procedure for estimation in contingency tables,” *The Annals of Mathematical Statistics*, 41, 907–917.
- (1972), “The analysis of incomplete multi-way contingency tables,” *Biometrics*, 28, 177–202.
- 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.
- Glas, Alexander, and Matthias Hartmann (2022), “Uncertainty measures from partially rounded probabilistic forecast surveys,” *Quantitative Economics*, 13, 972–1022.
- Ireland, C. T., and S. Kullback (1968), “Contingency tables with given marginals,” *Biometrika*, 55, 179–188.
- 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 (2026), “Survey design and professional forecasters: The case of uncertainty in the US SPF,” Discussion Paper 26-017, ZEW – Leibniz Centre for European Economic Research.

- Krüger, Fabian, Todd E. Clark, and Francesco Ravazzolo (2017), “Using entropic tilting to combine BVAR forecasts with external nowcasts,” *Journal of Business & Economic Statistics*, 35, 470–485.
- Krüger, Fabian, and Lora Pavlova (2024), “Quantifying subjective uncertainty in survey expectations,” *International Journal of Forecasting*, 40, 796–810.
- Lewis, Kurt F., and Charles H. Whiteman (2015), “Empirical Bayesian density forecasting in Iowa and shrinkage for the Monte Carlo era,” *Journal of Forecasting*, 34, 15–35.
- Lipsitz, Stuart R., and L. P. Zhao (1994), “Estimation in contingency tables with given marginals,” *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43, 223–230.
- 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.

Acknowledgements

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), the 2026 SNDE meeting (Lisbon), 2026 IAAE meeting (Lisbon), ECB, Boston University, and University of Pittsburgh. A supplementary online appendix is available [here](#).

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.

Todd E. Clark

Federal Reserve Bank of Cleveland, Cleveland, United States; Johns Hopkins University, Baltimore, Maryland, United States

Elmar Mertens (corresponding author)

European Central Bank, Frankfurt am Main, Germany; email: elmar.mertens@ecb.europa.eu

© European Central Bank, 2026

Postal address 60640 Frankfurt am Main, Germany

Telephone +49 69 1344 0

Website www.ecb.europa.eu

All rights reserved. Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the authors.

This paper can be downloaded without charge from www.ecb.europa.eu, from the [Social Science Research Network electronic library](#) or from [RePEc: Research Papers in Economics](#). Information on all of the papers published in the ECB Working Paper Series can be found on the [ECB's website](#).

PDF

ISBN 978-92-899-8016-6

ISSN 1725-2806

doi:10.2866/3181189

QB-01-26-221-EN-N