INTERNATIONAL MONETARY FUND

Scenario Synthesis and Macroeconomic Risk

Tobias Adrian, Domenico Giannone, Matteo Luciani, Mike West

WP/25/105

2025
MAY

WORKING PAPER

 WP/25/105

IMF Working Paper
Monetary and Capital Markets

Scenario Synthesis and Macroeconomic Risk

Prepared by Tobias Adrian, Domenico Giannone, Matteo Luciani, Mike West

Authorized for distribution by Domenico Giannone
May 2025

ABSTRACT: We develop methodology to bridge scenario analysis and risk forecasting, leveraging their respective strengths in policy settings. The methodology, rooted in Bayesian analysis, addresses the fundamental challenge of reconciling judgmental narrative approaches with statistical forecasting. Analysis evaluates explicit measures of concordance of scenarios with a reference forecasting model, delivers predictive synthesis of the scenarios to best match that reference, and addresses scenario set incompleteness. This provides a framework to systematically evaluate and integrate risks from different scenarios, aiding forecasting in policy institutions while supporting clear and rigorous communication of evolving risks. We also discuss broader questions of integrating judgmental information with statistical model-based forecasts in the face of as-yet unmodeled circumstance

JEL Classification Numbers:	C53, E27
Keywords:	Macroeconomic Forecasting; Mixtures of Scenarios, Misclassification Rates, Entropic Tilting, Bayesian Predictive Synthesis, Judgmental Forecasting, Forecast Risk Assessment
Author's E-Mail Address:	dgiannone@imf.org, tadrian@imf.org

* "The author(s) would like to thank" footnote, as applicable.

WORKING PAPERS

Scenario Synthesis and Macroeconomic Risk

Prepared by Tobias Adrian[1], Domenico Giannone, Matteo Luciani, Mike West

[1]Tobias Adrian, Director of the Monetary and Capital Markets Department, International Monetary Fund
700 19th Street NW, Washington, DC 20431, U.S.A.
tadrian@imf.org
[2]Domenico Giannone, Assistant Director, International Monetary Fund
700 19th Street NW, Washington, DC 20431, U.S.A.
dgiannon2@gmail.com
[3]Matteo Luciani, Principal Economist, Board of Governors of the Federal Reserve System
20th Street and Constitution Avenue NW, Washington, DC 20551, U.S.A.
matteo.luciani@frb.gov
[4]Mike West, The Arts & Sciences Distinguished Professor Emeritus of Statistics & Decision Sciences
Duke University, Durham, NC 27708, U.S.A.
mike.west@duke.edu

1 Introduction

Macroeconomic policy institutions such as central banks rely heavily on forecasting methods. Monetary policymakers are regularly briefed on the economic outlook, alternative policy paths, and the balance of risks around the central forecast. Central bank staff rely on a combination of structural macroeconomic models, reduced-form empirical models, and judgmental approaches to prepare such monetary policy briefings. The central forecast is used as a basis for alternative policy path discussions, and the balance of risks is discussed more loosely based on scenario analysis.

The Bank of England pioneered communication of risk with fan charts in 1993; the Inflation Reports show central projections of inflation with charts that reflect uncertainty. Uncertainty intervals are derived from judgmental assessments of risk around the baseline (Britton et al., 1998). Since 1995, the U.S. Federal Reserve's Tealbook (TB) has presented scenarios as perturbations around baseline forecasts. Most major central banks now use some variant of these approaches. Fan charts and scenario analysis pose practical challenges as they require frequent updating and quantification of risks based on judgment. Hence central banks are relying more often on statistical methods to forecast macroeconomic risk. The density forecasting approach of "Growth-at-Risk" (GaR: Adrian et al., 2016, 2019; Plagborg-Moller et al., 2020; Adrian et al., 2022) is increasingly popular. The Tealbook has included GaR measures together with scenarios since 2017; other central banks have also implemented GaR approaches in addition to the more judgmental scenario methods (e.g., Figueres and Jarociński, 2020; Lenza et al., 2023; Aikman et al., 2019; Eguren-Martin et al., 2024; Anesti et al., 2023; Jondeau et al., 2022; Alessandri et al., 2019).

Our focus here is on a formal statistical approach to integrating scenario-based balance of risk discussions with statistical forecasts. The methodology defines a synthesis of the baseline and scenarios that best match the statistical *reference* forecast distribution, the latter typically from GaR and/or a quantile regression model. The scenario synthesis assigns weights to each scenario, quantifying their relative concordance with the reference and so providing explication of why a certain set of scenarios is particularly relevant. The analysis also incorporates a synthetic "backstop" scenario designed to address potential incompleteness of the defined scenario set. In practice, uncertainty measures are usually published only for the baseline; alternative scenarios are typically represented only in terms of point forecasts. We use extensions of the Bayesian decision analysis method of entropic tilting (e.g. Robertson et al., 2005; Tallman and West, 2022) to define full scenario forecast distributions as perturbations the baseline.

Analysis further addresses scenario information beyond a single point forecast, specifically to use scenario tail percentiles that reflect measures of scenario risk. This links to the desirability of scenario hypotheses that represent more radical perturbations of the baseline than has been typical (e.g., Justiniano and Primiceri, 2008; Fernández-Villaverde et al., 2011; Adrian and Boyarchenko, 2012; He and Krishnamurthy, 2012; Brunnermeier and Sannikov, 2014; Fernández-Villaverde et al., 2023, 2024 with structural models, and Adrian et al., 2021; Caldara et al., 2021; Carriero et al., 2024 with reduced-form models). That scenarios considered by policy institutions often represent only modest perturbations of the baseline is also partly addressed by our use of the synthetic backstop scenario. This can serve as a "red flag" when the scenario set fails to account for risks– especially tail risks– supported under the statistical reference.

Relating baseline and scenario forecasts to the statistical reference exploits Bayesian predictive synthesis (e.g. McAlinn and West, 2019; Johnson and West, 2025) to motivate a discrete mixture (linear pool) of baseline and scenario distributions as a proxy for scenario-based forecasting.

The "match" of such a mixture with the statistical reference distribution uses a central measure of concordance between distributions, namely the *expected misclassification rate* (EMR). Identifying mixture weights to optimize EMR is then a formal Bayesian decision problem. Relative scenario probabilities based on this scenario:reference optimization guide evaluation and interpretation of the roles of scenarios. The analysis includes explicit statistical measures of *scenario set incompleteness* reflecting aspects of lack-of-concordance the scenario synthesis with the reference. This aids in the policy setting on the question of whether the baseline and chosen scenarios adequately reflect all the risks captured by the reference.

The case study draws on published versions of the TB. We use data from reports prepared for the December FOMC meetings in 2007 and 2018, giving predictions for 2008 and 2019, respectively. Following the TB, we focus on risks to real growth. Reanalysis incorporating other variables, such as inflation and unemployment at risk (Adams et al., 2021; Kiley, 2022; López-Salido and Loria, 2024) is straightforward but beyond our main scope here. Our detailed examples highlight the generation of scenario weights reflecting aspects of concordance with the reference, and also the questions of scenario set incompleteness. One example of that latter highlights the lack of a very negative, "downside risk scenario" in both the 2007 and 2018 TB. This relates to the particular interest in our analysis when economic uncertainty is high so that defining an adequate baseline forecast is challenging. Then, listing and discussing a range of plausible scenarios, each with an assigned probability derived from the reference match, offers a richer perspective on informed decision-making under uncertainty.

Our analysis takes baseline and scenarios (as well as the reference) as given. In policy practice, of course, the back-and-forth between changes to statistical forecast distributions and the evolving narrative of scenarios provides a rich ground to rigorously examine shifts in the balance of risks. This was noted by Bernanke (2023) and is germane to the TB, where scenario-based approaches to the balance of risk and statistical forecast distributions are discussed separately. As Federal Reserve Chair Jerome Powell noted during the Press Conference following the January 2025 FOMC meeting, *"One of the things our staff does is they look at a range of possible outcomes. [...] There'll be baseline, and then they'll show six or seven alternative scenarios, including really good ones and not so good ones. And what those do is they spark [...] the policymakers to sort of think and understand about [...] the uncertainties that surround us."* Our methodology provides formal cross-talk that can aid macroeconomic staff in policy institutions: it combines the communicative strength of narrative scenarios with the statistical rigor of predictive models, identifying the most relevant risks with easy-to-understand stories and quantifying the relevance of these stories.

Section 2 discusses foundations and overviews methodology. Section 3 addresses partial scenario information. Section 4 introduces expected misclassification rates as distributional concordance metrics, with foundational insights. Section 5 develops the embedding of scenario analysis in a fully Bayesian framework, with core theoretical summaries and aspects of computational implementation. Section 6 summarizes key aspects of the detailed case study. Section 7 links to broader questions of combining judgmental information with statistical model-based forecasts. The Appendix adds technical and methodological details. Summary comments define Section 8.

2 Setting, Foundations and Perspective

2.1 Context and Goals

Interest lies in forecasting a vector outcome $\mathbf{y}$, such as a path of several macroeconomic indicators over multiple future time periods, based on the following ingredients.

- A policy-based analysis produces a predictive density $p_0(\mathbf{y})$, referred to as the *baseline density*.
- Relative to the baseline, the policy analysis considers each of a set of *alternative scenarios*; scenario j, labeled $\mathcal{S}_j$, generates a predictive density $p_j(\mathbf{y})$. These are regarded as hypothetical scenarios to be assessed relative to the baseline.
- The baseline is a given forecast distribution in the policy setting, so not an hypothetical scenario; that understood, we use $\mathcal{S}_0$ and the index $j = 0$ to designate the baseline.
- Separately, a statistical model (e.g., the statistical GaR analysis) produces a full predictive density $p(\mathbf{y})$, referred to as the *reference predictive density*.

The over-arching goal is to identify "closeness" of each scenario to the reference $p(\mathbf{y})$, and rank them relative to that assessment. The methodology we introduce addresses this, building on foundational statistical concepts and model developments now discussed.

2.2 Scenario Mixtures and Bayesian Predictive Synthesis

A Bayesian decision-maker in the policy setting can regard the set of scenario p.d.f.s $p_j(\mathbf{y})$ as "information" to use in forming a policy-relevant overall forecast. This involves some form of *pooling* of the predictions across the baseline and alternative scenarios. Here the foundational theory of Bayesian predictive synthesis (BPS)– and the specific class of "mixture BPS" models (McAlinn and West, 2019, section 2.2; Johnson and West, 2025)– applies. Under BPS, a valid Bayesian predictive analysis can be based on a *scenario mixture*, i.e., a distribution with p.d.f. $f(\mathbf{y}|\boldsymbol{\alpha})$ that is a linear pool of the $p_j(\mathbf{y})$ with respect to probability weights α_j in a vector $\boldsymbol{\alpha}$, namely $f(\mathbf{y}|\boldsymbol{\alpha}) \propto \sum_j \alpha_j p_j(\mathbf{y})$.

A key theoretical aspect of mixture BPS is that it can address the broad question of "scenario set (in-)completeness". That is, a setting in which the baseline $\mathcal{S}_0$ and all of the the alternative scenarios $\mathcal{S}_j$ considered are discordant with the reference $p(\mathbf{y})$. This relates to the "model set incompleteness" issue widely discussed in Bayesian econometrics. BPS theory addresses this by requiring an additional p.d.f. to extend the initial set and to use in the mixture. This has been exploited in BPS applications– and in its generalization to decision-guided settings (BPDS: Tallman and West, 2023; Chernis et al., 2024)– by structuring the additional p.d.f. as a *backstop* that can be expected to be supported by future data that is not so well-predicted by the initial model set. Examples in the above studies use an over-dispersed average of the initial mixture of model p.d.f., and this strategy can be adopted for scenario analysis. The specific construction of such a backstop scenario in the case study in Section 6 provides an example of this modelling strategy.

Index the alternative scenarios from the policy setting by $j = 1 : J-1$ with the baseline $j = 0$ and now with $j = J$ for the chosen backstop p.d.f. The latter is labeled $\mathcal{S}_J$ though it is a purely synthetic scenario chosen for the above purposes. Then the overall *scenario mixture* p.d.f. is

$$f(\mathbf{y}|\boldsymbol{\alpha}) = \sum_{j=0:J} \alpha_j p_j(\mathbf{y}). \tag{1}$$

2.3 Incomplete Specification of Scenario Forecast Distributions

Scenario p.d.f.s $p_j(\mathbf{y})$ are typically only partially specified. A common setting is that $\mathcal{S}_j$ defines point forecasts such as means or medians, with or without uncertainty measures such as a few other percentiles. The foundational concept is that the alternative scenarios represent economically relevant "what-if?" perturbations of the baseline. Hence receiving such partial information on $\mathcal{S}_j$ indicates a modification of $p_0(\mathbf{y})$ to match that partial information. Our approach aims to identify $p_j(\mathbf{y})$ that is "closest to" the baseline $p_0(\mathbf{y})$ subject to being consistent with that partial scenario information. The theoretical basis for methodology to do this, detailed in Section 3.2, is that of entropic tilting (ET: Tallman and West, 2022). Since its introduction by Robertson et al. (2005), ET–based methodology has seen increasing use in forecasting in econometrics, finance and related areas (e.g. Krüger et al., 2017; Metaxoglou et al., 2018; Koop et al., 2019; Antolín-Díaz et al., 2021; Clark et al., 2022; West, 2024; Crump et al., 2025). The current setting is different, though use here of ET is close in spirit and goals to its original use in imposing constraints on a given– here the baseline– forecast distribution.

2.4 Scenario-Reference Concordance

The goal of measuring concordance of scenarios with the statistical reference is now that of relating $f(\mathbf{y}|\boldsymbol{\alpha})$ in eqn. (1) to the reference p.d.f. $p(\mathbf{y})$. This is addressed by identifying the probability vector $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_J)'$ such that the scenario mixture is "closest to" $p(\mathbf{y})$. This requires specification of a utility function to characterize and quantify "close" in comparing densities, and then the resulting methodology to evaluate $\boldsymbol{\alpha}$ and thus define both scenario-specific weights and the overall mixture synthesis. Section 4.1 introduces a foundational metric for this– based on a measure of concordance of $f(\mathbf{y}|\boldsymbol{\alpha})$ and $p(\mathbf{y})$ from traditional statistical classification. With some new and relevant theoretical results and motivating examples, this underlies its use in scenario synthesis.

3 Partial Scenario Information and Entropic Tilting

3.1 Partial Scenario Information

As noted in Section 2.3 the common setting is that for each scenario only partial information relative to the fully specified baseline is provided. In many examples, the partial information can be represented as expectations of functions of $\mathbf{y}$, and this is the setting we adopt. Often, only the perturbed central tendency is reported. If taken as a mean, it would be a constraint on the expected value directly. If taken as the median, then it is formally defined as the expectation of an indicator function. Similar reasoning applies to other percentiles. Our analysis below addresses multiple scenario features simultaneously, such as a set of percentiles.

Suppose that $\mathcal{S}_j$ provides partial information on $p_j(\mathbf{y})$ in terms of $\mathbf{m}_j = \mathrm{E}[\mathbf{s}_j(\mathbf{y})|\mathcal{S}_j]$ where $\mathbf{s}_j(\mathbf{y})$ is a q_j-vector of *scenario scores*; call the given vector $\mathbf{m}_j$ the *target score* for $\mathcal{S}_j$. In general, the definition of scores can be scenario-specific, but here we assume that $q_j = q$ and $\mathbf{s}_j(\mathbf{y}) = \mathbf{s}(\mathbf{y})$ for all $j = 1{:}J$. A main case of interest has elements of $\mathbf{s}(\mathbf{y})$ as indicator functions in one or more of the univariate dimensions; then $\mathbf{m}_j$ is a given vector of percentiles of $p_j(\mathbf{y})$ in those dimensions. With $\mathcal{S}_j$ regarded as a perturbations of the baseline, methodology aims at identifying that $p_j(\mathbf{y})$ closest to the baseline $p_0(\mathbf{y})$ subject to being consistent with the forecast information $\mathbf{m}_j$. Entropic tilting (ET) results if we choose to define "close to" in a Küllback-Leibler (KL) sense.

3.2 Entropic Tilting and Scenario-Baseline ET Weights

ET–based methodology, recently exploited in new ways in Bayesian predictive decision synthesis (e.g. Chernis et al., 2024; Tallman and West, 2023, 2025), was originally used in imposing constraints on forecast distributions; that is the context here. In our setting, ET aims to identify $p_j(\mathbf{y})$ to minimize the KL divergence of the baseline $p_0(\mathbf{y})$ from $p_j(\mathbf{y})$ subject to $\mathbf{m}_j = \mathrm{E}[\mathbf{s}_j(\mathbf{y})|\mathcal{S}_j] = \int_{\mathbf{y}} \mathbf{s}_j(\mathbf{y}) p_j(\mathbf{y}) d\mathbf{y}$. ET theory (Tallman and West, 2022) yields

$$p_j(\mathbf{y}) = k_j \mathrm{e}^{\boldsymbol{\tau}_j' \mathbf{s}_j(\mathbf{y})} p_0(\mathbf{y}) \quad \text{where} \quad k_j^{-1} = \int_{\mathbf{y}} \mathrm{e}^{\boldsymbol{\tau}_j' \mathbf{s}_j(\mathbf{y})} p_0(\mathbf{y}) d\mathbf{y}, \tag{2}$$

in which $\boldsymbol{\tau}_j$ is the (provably unique) *tilting vector* such that the expectation constraint is satisfied. The implied identity $\mathbf{0} = \int_{\mathbf{y}} \{\mathbf{s}_j(\mathbf{y}) - \mathbf{m}_j\} \exp\{\boldsymbol{\tau}_j' \mathbf{s}_j(\mathbf{y})\} p_0(\mathbf{y}) d\mathbf{y}$ is typically efficiently solved for $\boldsymbol{\tau}_j$ using simple Newton-Raphson.

In practice, it is typical that the baseline is represented in terms of a Monte Carlo (MC) sample, i.e., defined as a discrete distribution $\{\mathbf{y}^i, w_0^i\}_{i=1:n}$ with support points $\mathbf{y}^i$ having weight (probability) w_0^i. This is particularly key in our setting as we will later use importance sampling to evaluate $p_0(\mathbf{y})$ relative to the statistical reference $p(\mathbf{y})$. Then expectations defining the ET tilting vectors $\boldsymbol{\tau}_j$ are trivially evaluated via simple Monte Carlo integration.

ET analysis can be regarded as using $p_0(\mathbf{y})$ as an importance sampling proposal with respect to a target p.d.f. $p_j(\mathbf{y})$. This was recognized by Robertson et al. (2005) and provides useful numerical checks on consistency of the scenario-specific moment constraints with the baseline. On sample values $\mathbf{y}^i$, the implied normalized IS weights for MC integration in eqn. (2) are $w_j^i \propto u_j^i w_0^i$ where $u_j^i \propto p_j(\mathbf{y}^i)/p_0(\mathbf{y}^i) = \exp\{\boldsymbol{\tau}_j' \mathbf{s}_j(\mathbf{y}^i)\}$. The u_j^i are called *ET weights*. The standard expected sample size (ESS) can be evaluated on the u_i. ESS– the reciprocal of the sum of squared u_j^i over $i = 1{:}n$– provides an overall assessment of concordance of the $\mathcal{S}_j$ constraints with the baseline (Tallman and West, 2022, sect. 1.6). This relates closely to the minimized KL divergence (e.g., Gruber and West, 2016, sect. 3.3; Gruber and West, 2017, sect. 5.4) but on an interpretable scale.

4 Predictive Concordance

4.1 Predictive Concordance and Misclassification Rates

Predictive concordance mooted in Section 2.4 is presented here in a general setting comparing two density functions $p(\mathbf{y})$ and $f(\mathbf{y})$. The scenario mixture setting then arises with $f(\mathbf{y})$ replaced by $f(\mathbf{y}|\boldsymbol{\alpha})$ of eqn. (1) for any given $\boldsymbol{\alpha}$. Assume that $p(\mathbf{y})$ and $f(\mathbf{y})$ have the same support.

Suppose a random draw $\mathbf{y}$ is made from either $f(\mathbf{y})$ or $p(\mathbf{y})$ with equal probabilities. It is not disclosed which distribution generates the outcome $\mathbf{y}$. Write $\mathcal{H}_p$ for the hypothesis that $\mathbf{y} \sim p(\mathbf{y})$, and $\mathcal{H}_f$ for the hypothesis that $\mathbf{y} \sim f(\mathbf{y})$. Since the choice is made with $\Pr(\mathcal{H}_p) = \Pr(\mathcal{H}_f) = 0.5$, the resulting posterior probabilities conditional on the observed $\mathbf{y}$ are $\mathrm{P}(\mathcal{H}_p|\mathbf{y}) = p(\mathbf{y})/\{p(\mathbf{y}) + f(\mathbf{y})\}$ and $\Pr(\mathcal{H}_f|\mathbf{y}) = 1 - \mathrm{P}(\mathcal{H}_p|\mathbf{y})$.

Now assume that $\mathbf{y}$ is actually a draw from $p(\mathbf{y})$, i.e., condition on $\mathcal{H}_p$. Before learning $\mathbf{y}$, the expected posterior probability on $\mathcal{H}_f$ is then

$$\pi_{pf} \equiv \mathrm{E}[\mathrm{P}(\mathcal{H}_f|\mathbf{y})|\mathcal{H}_p] = \int_{\mathbf{y}} \mathrm{P}(\mathcal{H}_f|\mathbf{y}) p(\mathbf{y}) d\mathbf{y} = \int_{\mathbf{y}} \frac{f(\mathbf{y})p(\mathbf{y})}{\{f(\mathbf{y}) + p(\mathbf{y})\}} d\mathbf{y}. \tag{3}$$

By symmetry, if $\mathbf{y}$ is actually from $\mathcal{H}_f$, the expected posterior probability $\pi_{fp} = \mathrm{E}[\mathrm{P}(\mathcal{H}_p|\mathbf{y})|\mathcal{H}_f]$ is obviously the same, $\pi_{fp} = \pi_{pf}$.

Predictive concordance of $f(\mathbf{y})$ with $p(\mathbf{y})$ is inherently measured by the *expected misclassification rate (EMR)* π_{pf}. Higher values indicate that it is difficult to discriminate $f(\mathbf{y})$ from $p(\mathbf{y})$– indicating that draws from $f(\mathbf{y})$ are more likely to be misclassified as coming from $p(\mathbf{y})$– and vice-versa. This is a natural, interpretable metric to assess concordance– or discordance– of the two distributions.

In traditional classification in statistics and machine learning, the optimal Bayesian classifier judges $\mathbf{y}$ as coming from $f(\mathbf{y})$ with probability $\mathrm{P}(\mathcal{H}_f|\mathbf{y})$. Averaging across $\mathbf{y} \sim p(\cdot)$, and using standard terminology, $1 - \pi_{pf}$ is then both the population *sensitivity* and (due to the comparison of just two distributions and the implied symmetry) the population *sensitivity* of the optimal Bayesian classifier. It follows that $1-\pi_{pf}$ is the traditional overall *accuracy* of the test comparing $f(\cdot)$ and $p(\cdot)$, and so EMR $\pi_{pf} = 1 -$ *accuracy* is the traditional *error rate*. Increasing EMR indicates decreased discrimination of $f(\cdot)$ from $p(\cdot)$. Judging $f(\cdot)$ to be "close to" $p(\cdot)$ at higher values of π_{pf} is thus theoretically fundamental and practically interpretable.

It is immediate that $\pi_{pf} \leq 0.5$ with equality only when $f(\cdot) \equiv p(\cdot)$, defining the absolute scale for assessment of concordance. To prove this, note that $\pi_{pf} = \mathrm{E}[r(\mathbf{y})/\{1 + r(\mathbf{y})\}|\mathcal{H}_p]$ where $r(\mathbf{y}) = f(\mathbf{y})/p(\mathbf{y})$ with $\mathrm{E}[r(\mathbf{y})|\mathcal{H}_p] = 1$. Now, $r/(1 + r)$ is concave on $r > 0$ so that $\pi_{pf} \leq \mathrm{E}[r(\mathbf{y})|\mathcal{H}_p]/\{1 + \mathrm{E}[r(\mathbf{y})|\mathcal{H}_p]\} = 1/2$. The upper bound is achieved when $f(\mathbf{y}) \equiv p(\mathbf{y})$, i.e., $r(\mathbf{y}) = 1$ for all $\mathbf{y}$.

Now consider a decision setting where $f(\cdot)$ is to be chosen to be "close to" $p(\mathbf{y})$, and when $\mathbf{y} \sim p(\cdot)$. Choosing $f(\cdot)$ to maximize π_{pf} subject to relevant constraints is the optimal decision with respect to the implied constrained version of utility function $\mathrm{P}(\mathcal{H}_f|\mathbf{y})$. This defines the Bayesian foundation of use of EMR in the scenario synthesis development in Section 5.

4.2 Relationships to Küllback-Leibler Divergence

Note that $\pi_{pf} = \mathrm{E}[1/[1 + \exp\{k(\mathbf{y})\}]|\mathcal{H}_p]$ where $k(\mathbf{y}) = \log\{p(\mathbf{y})/f(\mathbf{y})\}$. Under $\mathcal{H}_p$, the scalar random quantity $k(\mathbf{y})$ has expectation $\mathrm{KL}(p\|f) \equiv \mathrm{E}[k(\mathbf{y})|\mathcal{H}_p] = \int_y \log\{p(\mathbf{y})/f(\mathbf{y})\}p(\mathbf{y})d\mathbf{y}$, the Küllback-Leibler divergence *of* $f(\cdot)$ *from* $p(\cdot)$. Assuming this expectation is finite, the delta approximation yields $\pi_{pf} \approx 1/[1 + \exp\{\mathrm{KL}(p\|f)\}]$; thus choosing $f(\cdot)$ to maximize π_{pf} is approximately the KL divergence minimizing solution. In many cases of practical relevance, this also provides a strict lower bound on π_{pf}, i.e., $\pi_{pf} \geq 1/[1 + \exp\{\mathrm{KL}(p\|f)\}]$; see Appendix A.1. Both the direct approximation and the lower bound are accurate in cases of higher concordance. Then, the symmetry of EMR in $f(\cdot)$ and $p(\cdot)$ implies that the same results hold with the two densities exchanged. With $\mathrm{KL}(f\|p)$ the divergence of *of* $p(\cdot)$ *from* $f(\cdot)$, this immediately refines the lower bound to $\pi_{pf} \geq 1/[1 + \exp(\kappa_{pf})]$ where $\kappa_{pf} = \min\{\mathrm{KL}(p\|f), \mathrm{KL}(f\|p)\}$, with equality as the direct delta approximation. KL divergence always raise the question of directional definition. This does not arise in using π_{pf} due to its symmetry, and this link to KL indicates the relevant "symmetrization" of KL as κ_{pf}. In cases of relatively good concordance, the two directional measures will also be close. Additional aspects of the relationship are discussed and exemplified in Appendix A.

EMR is fundamental for reasons discussed above; we have presented these connections to KL as it is a well-known measure. A major caveat is that it assumes KL measures are finite. There are important practical contexts where this is not so. An example has $f(\mathbf{y})$ Gaussian and $p(\mathbf{y})$ log T with any degrees of freedom; then $p(\mathbf{y})$ has no moments at all (e.g. West, 2024, Supplementary Material, Appendix B) and $\mathrm{KL}(p\|f)$ is infinite. In contrast, $\pi_{pf} \in (0, 0.5]$ always.

5 Scenario Synthesis

5.1 EMR and Optimizing Scenario Mixture Probabilities

The predictive concordance concept applies to the scenario mixture setting with $f(\mathbf{y})$ replaced by $f(\mathbf{y}|\boldsymbol{\alpha}) = \sum_{j=0:J} \alpha_j p_j(\mathbf{y})$ at any chosen probability vector $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_J)'$. Making dependence on $\boldsymbol{\alpha}$ explicit, eqn. (3) is now

$$\pi_{pf}(\boldsymbol{\alpha}) = \int_{\mathbf{y}} \frac{f(\mathbf{y}|\boldsymbol{\alpha})p(\mathbf{y})}{\{f(\mathbf{y}|\boldsymbol{\alpha}) + p(\mathbf{y})\}} d\mathbf{y}. \tag{4}$$

Values of $\boldsymbol{\alpha}$ yielding high values of $\pi_{pf}(\boldsymbol{\alpha})$ define mixtures of baseline and scenarios "close to" to the reference $p(\mathbf{y})$ in terms of probabilistic concordance. Suppose $\widehat{\boldsymbol{\alpha}}$ maximizes $\pi_{pf}(\boldsymbol{\alpha})$ with maximum value $\widehat{\pi}_{pf} = \pi_{pf}(\widehat{\boldsymbol{\alpha}})$. Each element $\widehat{\alpha}_j$ of $\widehat{\boldsymbol{\alpha}}$ quantifies the extent to which $\mathcal{S}_j$ is concordant with the reference *relative to the other $\mathcal{S}_i$ for $i \neq j$*. The summary $\widehat{\pi}_{pf}$ is a concrete measure of the concordance of the set of predictions from the baseline and the scenarios combined. A low value of $\widehat{\pi}_{pf}$ indicates that none of the $\mathcal{S}_j$ nor their mixture are really concordant with the reference, relating to the scenario set incompleteness discussion of Section 2.2. Thus $\widehat{\pi}_{pf}$ measures how "discordant" the scenario set is with the reference statistical predictions. The weight $\widehat{\alpha}_J$ on the backstop provides additional information.

The framework addresses selection of $\boldsymbol{\alpha}$ as a decision problem that maximizes $\pi_{pf}(\boldsymbol{\alpha})$ with regularization to penalize very small α_j. This is based on deeper foundational and theoretical development in the next subsection, and leads to choosing $\boldsymbol{\alpha}^*$ to maximize the objective function

$$\lambda(\boldsymbol{\alpha}) = \log\{\pi_{pf}(\boldsymbol{\alpha})\} + \epsilon \sum_{j=0:J} \log(\alpha_j), \quad \text{subject to } \alpha_j > 0 \ (j = 0{:}J) \quad \text{and} \sum_{j=0:J} \alpha_j = 1, \tag{5}$$

where $\epsilon > 0$ is a very small regularization parameter. As we now show, eqn. (5) is in fact the log of a formal posterior distribution so that the optimization seeks the posterior mode.

5.2 Bayesian Foundation

5.2.1 EMR is a Likelihood Function

Suppose that the economic reality $\mathbf{y}$ is generated from the reference $p(\mathbf{y})$ and consider an hypothetical/synthetic binary outcome z generated from the Bernoulli distribution with success probability $\Pr(z = 1|\mathbf{y}, \boldsymbol{\alpha}) = f(\mathbf{y}|\boldsymbol{\alpha})/\{p(\mathbf{y}) + f(\mathbf{y}|\boldsymbol{\alpha})\}$. Then

$$p(z = 1, \mathbf{y}|\boldsymbol{\alpha}) = \Pr(z = 1|\mathbf{y}, \boldsymbol{\alpha})p(\mathbf{y}|\boldsymbol{\alpha}) = \Pr(z = 1|\mathbf{y}, \boldsymbol{\alpha})p(\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\alpha})p(\mathbf{y})/\{p(\mathbf{y}) + f(\mathbf{y}|\boldsymbol{\alpha})\}.$$

Now suppose you observe $z = 1$ but not $\mathbf{y}$; EMR emerges via expectations over the "missing data" $\mathbf{y}$, viz., $p(z = 1|\boldsymbol{\alpha}) = \pi_{pf}(\boldsymbol{\alpha})$. Thus, $\pi_{pf}(\boldsymbol{\alpha})$ is in fact a likelihood function for the *parameter* $\boldsymbol{\alpha}$ based on an hypothetical observation $z = 1$ that classifies a random draw from $p(\mathbf{y})$ as coming from $f(\mathbf{y})$ under a 50:50 prior. The connection with the foundation of EMR in Section 4.1 is immediate.

It follows that EMR-maximizer $\widehat{\boldsymbol{\alpha}}$ is a maximum likelihood estimate (MLE). Evaluating $\widehat{\boldsymbol{\alpha}}$ is probability simplex constrained convex optimization problem with a unique solution, the convexity and hence uniqueness being shown here in Appendix C. The solution will typically be a *sparse mixture* of scenarios, with some zeros in $\widehat{\boldsymbol{\alpha}}$. This follows from general results of optimization of

convex functions over the probability simplex (e.g. Boyd and Vandenberghe, 2004). For some integer $k \in \{0 : J\}$ a subset of k of the $\widehat{\alpha}_j$ can be zero. There are cases when $k = 0$ but $k > 0$– defining a sparse optimizing vector– is more usual, especially with larger J and diversity among the $p_j(\mathbf{y})$. This relates to general features of optimization over the simplex; simplex constraints operate to shrink weights to the boundaries, effectively as ℓ_1 shrinkage for sparsity (e.g. Brodie et al., 2009). This underlies the notion of scalability of the analysis to larger numbers of scenarios.

However, sparsity in $\widehat{\boldsymbol{\alpha}}$ is unstable since it is not a genuine feature but is induced by the implicit prior ℓ_1 penalty; its values are typically very sensitive to small changes in the input scenario and reference p.d.f.s. This pathology of sparsity inducing penalties was identified and documented in the context of forecasting by Giannone et al. (2021). In the current setting, take an example with two very similar scenario p.d.f.s; one of these scenarios will have a zero value in $\widehat{\boldsymbol{\alpha}}$, the other non-zero. Then, a very small change in either of the p.d.f.s– or of the reference p.d.f.– will flip the zero/non-zero pattern. At each of these extremes– and for ranges of the α_j on these two scenarios bridging the extremes– the resulting scenario mixture $f(\mathbf{y}|\widehat{\boldsymbol{\alpha}})$ will be almost unchanged. This sensitivity is undesirable; it is desirable to have similar probabilities on the two scenarios. The key point is that a uniform prior on $\boldsymbol{\alpha}$ favors overly sparse models when the likelihood function has modes at the simplex boundaries. This can be addressed by imposing additional constraints or, more foundationally, with a minimally informative "regularizing" prior over $\boldsymbol{\alpha}$.

5.2.2 Priors and Penalties

The natural priors are Dirichlet, $\boldsymbol{\alpha} \sim \text{Dir}(\mathbf{a})$ having p.d.f. $p(\boldsymbol{\alpha}) \propto \prod_{j=0:J} \alpha_j^{a_j - 1}$ over the simplex. Here $a_j > 0$ for all j and, with precision $a = \sum_{j=0:J} a_j$, the means are a_j/a and prior joint mode has elements $\max\{0, (a_j - 1)/(a - J - 1)\}$. A prior with each $a_j = 1 + \epsilon$ for a very small $\epsilon > 0$ is "minimally informative" subject to the joint prior mode being positive on each scenario. Modifications to $a_j = 1 + \epsilon_j$ to differentially favor scenarios *a priori* are obviously of interest, but for this paper the symmetric prior is adopted. For given ϵ, the prior joint mode and mean are then each $\mathbf{1}/(J + 1)$, i.e., favoring a uniform set of scenario probabilities though with high uncertainty since ϵ is taken as very small. Under this prior $\boldsymbol{\alpha} \sim \text{Dir}(\mathbf{1}(1 + \epsilon))$, the log posterior is $\lambda(\boldsymbol{\alpha})$ in eqn. (5), up to an additive constant. The prior is zero at simplex boundaries, hence so is the resulting unimodal posterior. The posterior mode– denoted by $\boldsymbol{\alpha}^*$– maximizes EMR modified by the prior-based penalty that explicitly acts to move from the boundary zero MLE values in $\widehat{\boldsymbol{\alpha}}$ to small but non-zero values. This leads to more stable and robust results and addresses the issues discussed in the previous section.

Analysis requires choice of a (small) value of the regularizing hyper-parameter ϵ. Based on theory in Appendix B, the default recommendation is $\epsilon = c/(J+1)$, where $c = 0.005$. The value of c can be modified somewhat up/down with minimal impact, while the scaling with number of scenarios is important in more heavily penalizing the MLE-based analysis in higher dimensions. Given $\epsilon > 0$, evaluation of the posterior mode $\boldsymbol{\alpha}^*$ to maximize eqn. (5) trivially modifies the probability simplex constrained convex optimization problem with a unique solution. See Appendix C.

It is also of interest to consider analyses with additional constraints on $\boldsymbol{\alpha}$. A key example is to require $\alpha_0 \geq \alpha_j$ for $j = 1{:}J$, consistent with the view that the baseline is the "modal" scenario. In general it is of interest to run comparative analyses with and without such constraints. Such a constraint simply modifies the Dirichlet prior by the indicator of the constraint; this does not impact the convexity of the optimization problem and is trivially implemented.

5.3 Monte Carlo Importance Sampling

Analysis *prima facie* relies on evaluating the p.d.f.s $p(\mathbf{y})$ and each $p_j(\mathbf{y})$, and then performing the integration in eqn. (4). Analytic approximations to the integral may be explored. Specific approximations relate to measures of discriminatory information in classification using mixtures (e.g. Lin et al., 2016). In practice, however– and as already noted in Section 3.2– forecasts will typically be "available" in terms of Monte Carlo (MC) samples, so direct evaluation of π_{pf} by MC integration is a priority (and avoids concerns of assessing the quality of analytic approximations).

The analysis is implemented with the values of the p.d.f.s $p(\mathbf{y})$ and the $p_j(\mathbf{y})$ available only on a (large) sample of MC draws from the reference $p(\mathbf{y})$, a *reference random sample*. The random sample $\mathbf{y}^i$, $(i = 1{:}n)$, is drawn from $p(\mathbf{y})$ and at the first step this defines an importance sample (IS) for the baseline $p_0(\mathbf{y})$ with normalized IS weights $w_0^i \propto p_0(\mathbf{y}^i)/p(\mathbf{y}^i)$. The discrete distribution $\{\mathbf{y}^i, w_0^i\}_{i=1:n}$ defines the MC approximation to the baseline for evaluation of expectations in the downstream analysis. A proviso is that $p(\mathbf{y})$ is a relevant importance sampling proposal; in particular, it should be heavier-tailed than $p_0(\mathbf{y})$. As in all applications of IS, monitoring efficiency measures such as the % effective sample size ESS $= n^{-1}100/\sum_{i=1:n}(w_0^i)^2$ provides guidance; initial analysis generating a relatively low ESS guides choice of a larger sample size. This IS analysis then underlies evaluation of scenario-specific ET parameters as in Section 3.2, yielding ET weights $u_j^i \propto \exp\{\boldsymbol{\tau}_j'\mathbf{s}_j(\mathbf{y}^i)\}$ on sample $\mathbf{y}^i$ defining $p_j(\cdot)$ relative to the baseline. Scenario-specific ESS measures using the u_j^i weights are then relevant. In (rare) cases of a scenario that is really discordant with the baseline, a very low ESS indicates such. Refined but much more computationally demanding adaptive IS methods may be considered, but are outside our current scope. In any case, encountering such discordance would indicate that such a scenario might better be considered separately and its full distribution directly assessed.

This ET analysis leads to *compound weights* $w_j^i \propto u_j^i w_0^i$ relating $\mathcal{S}_j$ to the reference; these are called the *ET–IS weights*. ESS measures can now also be evaluated on the w_j^i to provide direct overall assessment of each $p_j(\cdot)$ relative to the reference $p(\cdot)$. Note that there can be cases where a scenario is more concordant with the reference than the baseline as some of our examples show. The reference sample and compound ET−IS weights are then ingredients in the direct evaluation of eqn. (4) via MC integration.

6 Case Study

The case study draws from the Risk and Uncertainty analyses in the December 2007 and 2018 Tealbooks (Federal Reserve Board, 2007, 2018). The scenarios specify point forecasts for GDP growth, inflation, the unemployment rate, and other variables. The methodology applies to multiple variables and horizons, but this first application restricts attention to one-year ahead GDP growth, namely $\mathbf{y} = y$, now scalar. Analysis follows the processes discussed in the previous sections; Appendix D gives a summary of the flow of analytic and computational details.

6.1 Reference Distribution

Among recent statistical approaches to risk assessment, Adrian et al. (2019) develop quantile regression models of conditional predictive distributions and show that financial markets provide useful risk information. This approach has influenced practice, being adopted for conditional one-year ahead forecasts of GDP growth, unemployment rate, and inflation, for example, by the

Federal Reserve Board (Engstrom and Gonzalez-Astudillo, 2017) in the "Time-Varying Macroeconomic Risk" exhibit in the Risk and Uncertainty section of Tealbook A, and the New York Federal Reserve (Boyarchenko et al., 2023) in Outlook-at-Risk. Other central banks and international financial institutions have also adopted GaR approaches (examples include Figueres and Jarociński, 2020; Lenza et al., 2023; IMF, 2017; Eguren-Martin et al., 2024; Anesti et al., 2023; Jondeau et al., 2022; Alessandri et al., 2019; Pujadas et al., 2022; Hafemann, 2023).

The Federal Reserve Board started producing the Tealbook Time-Varying Macroeconomic Risk forecasts in 2017 but do not provide past values. The NY Fed started producing the Outlook-at-Risk forecasts only in 2023 but provides past values starting in 1989. As a result, our case study constructs and compares reference distributions from both the NY Fed and the Fed Board. In practice, both the NY Fed and the Tealbook give five predictive percentiles (Table 1). To construct the reference distribution, we fit skew-t distributions (Azzalini and Capitanio, 2003) on these percentiles. Following Adrian et al. (2019), the four parameters of the skew-t minimize the squared distance between the given reference quantiles and those of the resulting skew-t. Table 3 reports resulting parameters.

Table 1: Reference percentiles

NY Fed			Tealbook	
	2007	2018		2018
P10:	−1.7	0.0	P5	0.7
P25:	0.2	1.1	P15	1.3
P50:	1.8	2.1	P50	2.5
P75:	3.3	3.0	P85	3.6
P90:	4.8	4.0	P95	4.3

The NY Fed Outlook at Risks gives point forecasts for one-year ahead GDP growth. The Tealbook Time-Varying Macroeconomic Risk gives one-year ahead Tealbook forecast errors in the December 2018 Tealbook, which provides GDP growth point forecasts once the baseline forecast is added.

6.2 Baseline Distribution

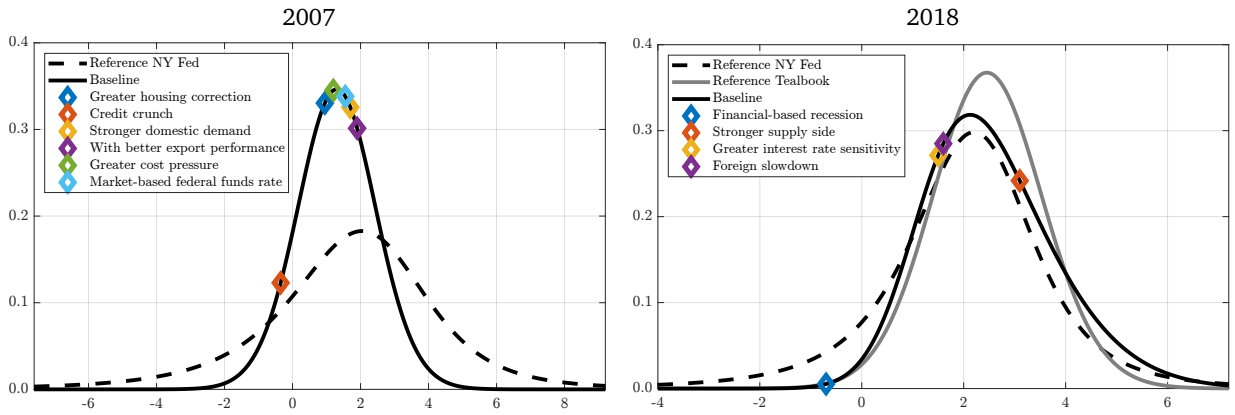
Table 2 shows baseline and alternative scenario projections. The Tealbook provides the point forecast and 70% intervals for the baseline. To construct the baseline distribution, we take the point forecast as the median and extremes of the 70% confidence interval as the 15th and 85th percentiles, respectively. We fit a skew-t with 50 degrees of freedom to these percentiles– the choice of degrees of freedom allows some modest tail-weight beyond normal but effectively represents a "close to normal" distribution.

Table 2: Tealbook Baseline and Scenarios

December 2007					December 2018			
j	Scenario $\mathcal{S}$	P50	P15	P85	j	Scenario $\mathcal{S}$	P50	P15	P85
0	Baseline	1.3	0.1	2.5	0	Baseline	2.4	1.2	3.9
1	Greater housing correction	1.0			1	Financial-based recession	−0.7		
2	Credit crunch	−0.4			2	Stronger supply side	3.1		
3	Stronger domestic demand	1.7			3	Greater interest rate sensitivity	1.5		
4	With better export performance	1.9			4	Foreign slowdown	1.6		
5	Greater cost pressure	1.2							
6	Market-based federal funds rate	1.6							

December 2007 TB: The baseline projection for 2008 is on page I-21, while the alternative scenarios are on page I-17—the scenarios for 2008 are obtained by averaging the values for 2008:H1 and 2008:H2. *December 2018 TB*: The baseline projection for 2019 is on page 88, while the alternative scenarios are on page 84.

Figure 1: Reference and baseline with scenario point forecasts



Solid black line is baseline p.d.f, and dashed black line is reference p.d.f. estimated from the NY Fed Outlook-at-Risk; solid gray line is the reference estimated from the Tealbook Time-Varying Macroeconomic Risk. Diamonds show point forecasts from scenarios.

Table 3: Reference and baseline skew-t parameters

	Distribution	Type	lc	sc	sk	df
2007	Reference	NY Fed	2.7	2.2	−0.5	3.4
	Baseline		1.3	1.1	0.0	50.0
2018	Reference	NY Fed	2.5	1.3	−0.3	3.0
	Reference	Tealbook	2.1	1.1	0.5	50.0
	Baseline		1.2	1.9	2.1	50.0

The skew-t parameters are those for location (lc), scale (sc), skewness (sk) and degrees of freedom (df).

Table 3 shows parameters of the baseline and reference skew-t distributions; Figure 1 shows the p.d.f.s. The baseline is much more precise than the NY Fed reference, with a lower scale and higher degrees of freedom. This raises questions for economic forecasting and policy design. In a world with a known "true model" of the economy, the baseline and reference would be identical; as they differ in practice, interpreting the scenarios is the challenge. By comparison, the baseline is roughly as precise as the Tealbook reference, the latter having 50 degrees of freedom as a result of the optimization process in fitting the skew-t. This suggests that the Tealbook reference may underestimate risk. As Federal Reserve Chair Jerome Powell said during the Press Conference following the January 2025 FOMC meeting, we should not be surprised as *"it is human nature, apparently, to underestimate [...] how fat the tails are.[...] We think of things in a normal distribution. And in the economy, it's not a normal distribution."*

Given reference and baseline distributions, our analysis proceeds based on Monte Carlo sampling from the reference. In developments below, the MC sample size is 10^6 and resulting MC analysis summaries stable and robust across reanalyses with such a large sample size.

6.3 Scenarios

TB scenarios in Table 2 provide only point forecasts. In the Dec. 2007 (2018) TB, we have 6 (4) alternative scenarios.[1] Figure 1 shows scenarios on the baseline, indicating their concentration around the baseline median but with some indication of downside risk in the left tail.

[1]We exclude "More room to grow" scenario in 2007, and both "Supply constraints" and "Lower oil prices" scenarios in 2018. These had GDP point forecasts identical to either the baseline or another scenario. If included in our synthesis, they would receive equal weight with either the baseline or one other scenario.

Our first analysis treats the scenario point forecasts as medians[2] of the $p_j(y)$, and the ET construction maps the baseline to each scenario p.d.f. constrained to its specified median only; see the P50s in Table 2. While the TB provides no measures of uncertainty around the scenario projections, such information could be useful and available in other applications. Section 6.5 explores analyses with P15 and P85 constructed for each scenario.

As discussed in Section 2.2, we augment the scenario set with a backstop located at the center of the scenarios while being relatively over-dispersed. Since the scenario information here is restricted to the median point forecasts, we first construct $p_j(y)$, $j = 1, \ldots, J$, using ET as in Section 3, then use the implied percentiles to define those of the backstop. Specifically, the backstop has $\text{P50}_\text{B} = \text{median}_{j=1,\ldots,J}\text{P50}_j$, $\text{P15}_\text{B} = \min_{j=1,\ldots,J}\text{P15}_j$, and $\text{P85}_\text{B} = \max_{j=1,\ldots,J}\text{P85}_j$, respectively. Numerical details are in Table 4.

Figure 2: Examples of Tilted Distributions of Alternative Scenarios
December 2007 Tealbook

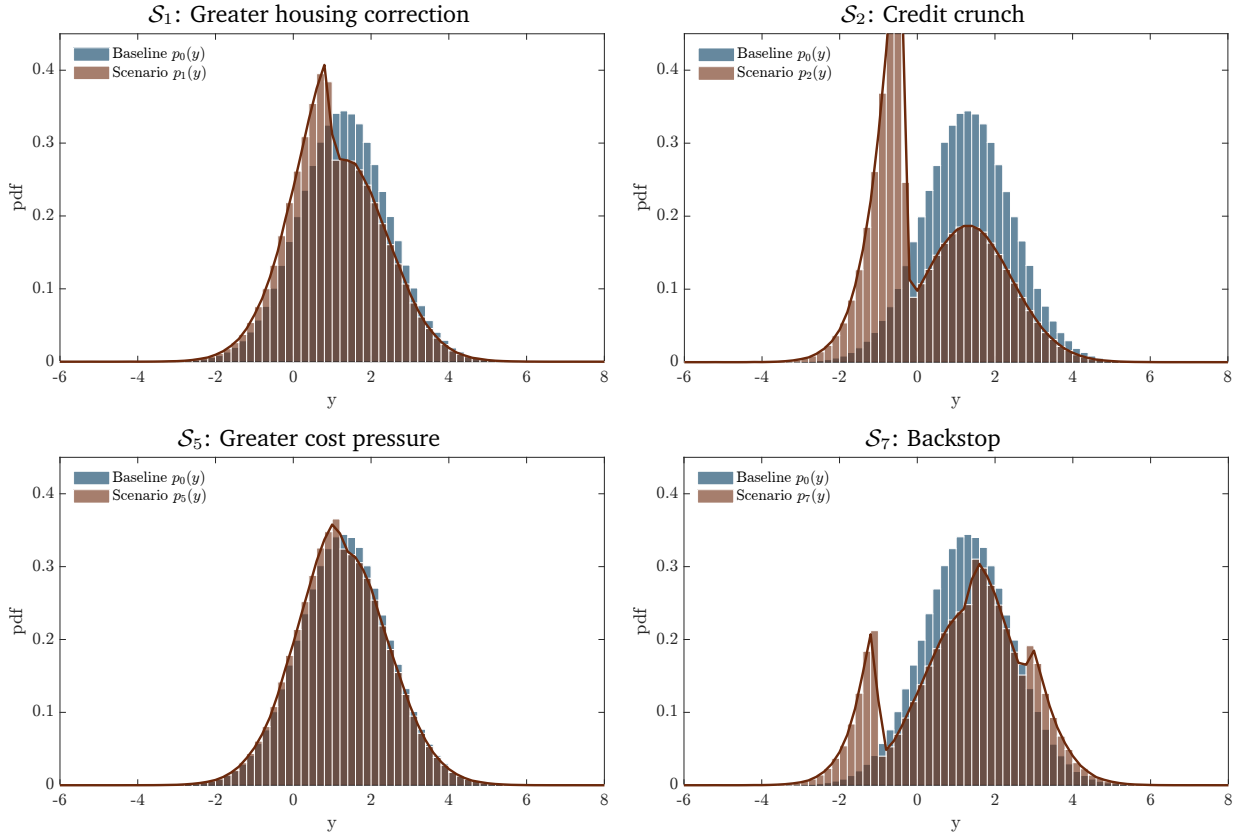


Figure 2 shows the resulting $p_j(y)$ for four of the 2007 TB scenarios. Table 4 shows corresponding ET–based ESS measures. When a scenario is close to the baseline (e.g., $\mathcal{S}_1$ and $\mathcal{S}_5$), the tilted distribution remains close and slightly asymmetric; otherwise, the tilted distribution can exhibit skewness and multimodality due to the mismatch between the scenario forecasts and baseline. The

[2]TB and other point forecasts might alternatively be treated as modes of scenario distributions. Our analysis can address that, based on new theoretical results (not reported here) showing why and how entropic tilting can be applied when point forecasts are modes. We use medians, however, based on fundamental concern for deeper representation of the probability distributions of scenarios, and embedding in more detailed analyses with multiple percentiles.

Table 4: Synthesis based on P50 information: NY Fed reference

TB	j	Scenario $\mathcal{S}$	P15$_j$	P50$_j$	P85$_j$	ET$_j$%	IS$_j$%	π_j^*	$\widehat{\boldsymbol{\alpha}}_j$	$\boldsymbol{\alpha}_j^*$
Dec. 2007	0	Baseline	0.1	1.3	2.5	100.0	62.6	0.41	0.31	0.27
	1	Greater housing correction	−0.1	0.9	2.3	94.3	57.4	0.40	0.00	0.02
	2	Credit crunch	−1.0	−0.4	2.0	28.7	30.9	0.36	0.08	0.08
	3	Stronger domestic demand	0.3	1.7	2.7	92.6	65.4	0.42	0.00	0.04
	4	Better export performance	0.4	1.9	2.9	84.3	65.2	0.42	0.31	0.27
	5	Greater cost pressure	0.0	1.2	2.5	99.5	61.3	0.41	0.00	0.02
	6	Market-based Fed Funds rate	0.2	1.6	2.6	97.1	64.8	0.41	0.00	0.03
	7	Backstop	−1.0	1.4	2.9	56.4	67.2	0.43	0.31	0.27
		$f(\mathbf{y}\|\widehat{\boldsymbol{\alpha}})$	−0.2	1.4	2.7		71.5	0.43		
		$f(\mathbf{y}\|\boldsymbol{\alpha}^*)$	−0.2	1.4	2.7		71.2	0.43		
Dec. 2018	0	Baseline	1.2	2.4	3.9	100.0	88.5	0.47	0.74	0.64
	1	Financial-based recession	−1.0	−0.6	3.1	0.6	8.4	0.35	0.04	0.04
	2	Stronger supply side	1.4	3.1	4.4	84.6	67.5	0.45	0.00	0.04
	3	Greater interest rate sensitivity	0.7	1.5	3.4	69.4	70.3	0.45	0.13	0.11
	4	Foreign slowdown	0.8	1.6	3.5	75.3	74.5	0.46	0.02	0.10
	5	Backstop	−1.0	1.6	4.4	2.1	37.9	0.43	0.07	0.08
		$f(\mathbf{y}\|\widehat{\boldsymbol{\alpha}})$	0.9	2.2	3.8		91.4	0.48		
		$f(\mathbf{y}\|\boldsymbol{\alpha}^*)$	0.9	2.2	3.8		90.9	0.48		

P15 and P85 for $\mathcal{S}_j$, $j = 1 : J-1$, are for the ET–based $p_j(y)$; the other percentiles are inputs to the analysis. ET$_j$ is the ET-based ESS of $p_j(y)$ relative to the baseline $p_0(y)$ while IS$_j$ denotes that for ET–IS implied relative to the reference $p(y)$. π_j^* is the EMR of the reference and scenario j alone. α_j^* and $\widetilde{\alpha}^*$ are the probabilities on $\mathcal{S}_j$ in the optimal mixture synthesis in analyses with and without the backstop scenario, respectively. The optimization constraints $\alpha_0 \geq \alpha_j, j = 1{:}J$, apply in both cases.

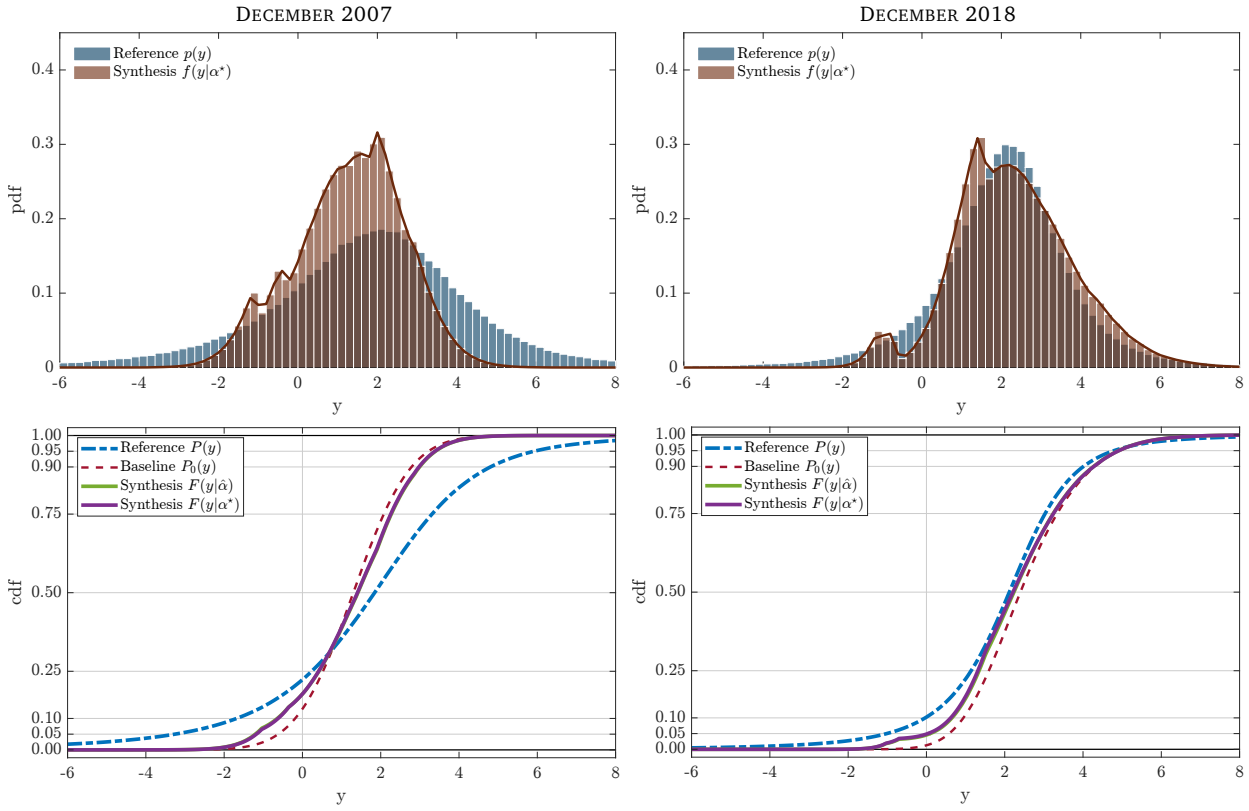
emergence of interesting shapes and multimodality in scenario distributions indicates the hypothesized state of the economy in the scenario is in regions poorly supported by the baseline. This is also related to the concept of modest policy intervention of Leeper and Zha (2003), i.e. that we can analyze policy effects using the baseline as long as entertained policy interventions are small enough that economic agents would not change their behavior in response to the intervention. Related considerations are those of formally down-weighting extreme conditional assumptions in conditional forecasting (e.g. Antolín-Díaz et al., 2021; Chernis et al., 2024, sect. 3).

6.4 Scenario Synthesis based on Medians

Table 4 shows optimal scenario mixture weights and other summaries from analyses. There is strong concordance between the mixture synthesis and the reference in the case of the 2018 TB (EMR=0.48), while the concordance is weaker in the 2007 Tealbook (EMR=0.43). Figure 3 provides insights via comparison of p.d.f.s. For TB 2008, the mixture synthesis is light-tailed relative to the reference, as all scenarios lack probability on downside GDP ranges that the reference meaningfully supports. For the 2018 example, the scenario mixture supports positive GDP values– partly as the baseline is already right skewed– but assigns relatively limited support to negative GDP growth.

In the 2007 TB analysis, the baseline, $\mathcal{S}_4$ and backstop are roughly equally weighted, followed by $\mathcal{S}_2$. The two more extreme scenarios $\mathcal{S}_2$ and $\mathcal{S}_4$ get weight as they help to capture the spread of the reference, while the other scenarios get small weights– they sit in the center of the reference

Figure 3: p.d.f. and c.d.f of scenario synthesis and NY Fed reference



Upper: p.d.f.s of the reference and scenario synthesis.*Lower:* c.d.f.s of the reference, baseline and scenario synthesis including results synthesis c.d.f.s based on both $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$; note that the latter are effectively indistinguishable.

and are very similar to the baseline, as shown by the ET ESS measures. In contrast, in the 2018 TB example the reference is more precise and the baseline is heavier weighted than other scenarios. Here the EMR of the baseline alone is quite high, so the alternative scenarios add small but limited value in contributing to approximating the reference.

A key feature of analysis is that $100-$ESS for the synthesis is an absolute measure of scenario set incompleteness relative to the reference. In TB 2007, the ESS of $f(\mathbf{y}|\boldsymbol{\alpha}^*)$ is about 71-72%; we can say that the scenario set is about 28-29% incomplete. In contrast, in TB 2018 the ESS of $f(\mathbf{y}|\boldsymbol{\alpha}^*)$ is about 91%, which indicates that the TB scenarios much more adequately represent the risk and uncertainty in the economy as defined by the reference than in 2007. A hint of the inability of TB 2007 scenarios to properly capture the risks in the reference comes also from the substantial weight on the backstop; in contrast, in TB 2018, the backstop receives low weight. As a general point looking ahead, a backstop p.d.f. that is over-dispersed has general benefits, but in application other choices are possible and may be preferred. These examples highlight this, indicating that scenarios reflecting increased support in the upper and lower tails of the– anticipated– reference distribution are most relevant. Future applications might address this.

Table 5: Synthesis based on P50 information: TB 2018, Tealbook reference

j	Scenario $\mathcal{S}$	P15_j	P50_j	P85_j	$\text{ET}_j\%$	$\text{IS}_j\%$	π_j^*	$\widehat{\alpha}_j$	α_j^*
0	Baseline	1.2	2.4	3.9	100.0	77.7	0.49	1.00	0.89
1	Financial-based recession	−1.0	−0.6	3.1	0.6	0.8	0.33	0.00	0.01
2	Stronger supply side	1.4	3.1	4.4	84.7	51.0	0.47	0.00	0.05
3	Greater interest rate sensitivity	0.7	1.5	3.4	69.6	56.3	0.44	0.00	0.02
4	Foreign slowdown	0.8	1.6	3.5	75.4	61.3	0.45	0.00	0.02
5	Backstop	−1.0	1.6	4.4	2.1	3.2	0.40	0.00	0.01
	$f(\mathbf{y}\|\widehat{\boldsymbol{\alpha}})$	1.2	2.4	3.9		77.7	0.49		
	$f(\mathbf{y}\|\boldsymbol{\alpha}^*)$	1.2	2.4	3.9		76.1	0.49		

Details as in Table 4.

Table 5 summarizes 2018 analysis using the reference distribution based on percentiles from the Tealbook Time-Varying Macroeconomic Risk to compare with the NY Fed-based details above. Again the synthesis here uses only the scenario medians. In this case, the baseline is as precise as the reference and has similar tail-weight in terms of the skew-t degrees of freedom. The main point, however, is that the baseline dominates the scenarios, indicating that the hypothesized median shifts they represent add little to no value in predictive discrimination relative to the reference.

On modeling strategy, consider an example of "normal" economics times as represented by 2018. In such settings, (i) the reference can be expected to be relatively light-tailed, (ii) scenarios can be expected to be modest in terms of varying from backstop, and (iii) the resulting scenario synthesis will be be close to the reference. There is limited scenario set incompleteness and the backstop will play a limited role. Contrast this with periods of higher uncertainty, such as in the 2007 context here where the reference distribution should have appropriately fatter tails. Then the synthesis of the baseline and the scenarios can substantially under-represent the reference unless the scenario set includes more extreme considerations. This mandates admitting extreme scenario considerations as a rational response to increased uncertainty in very uncertain economic times.

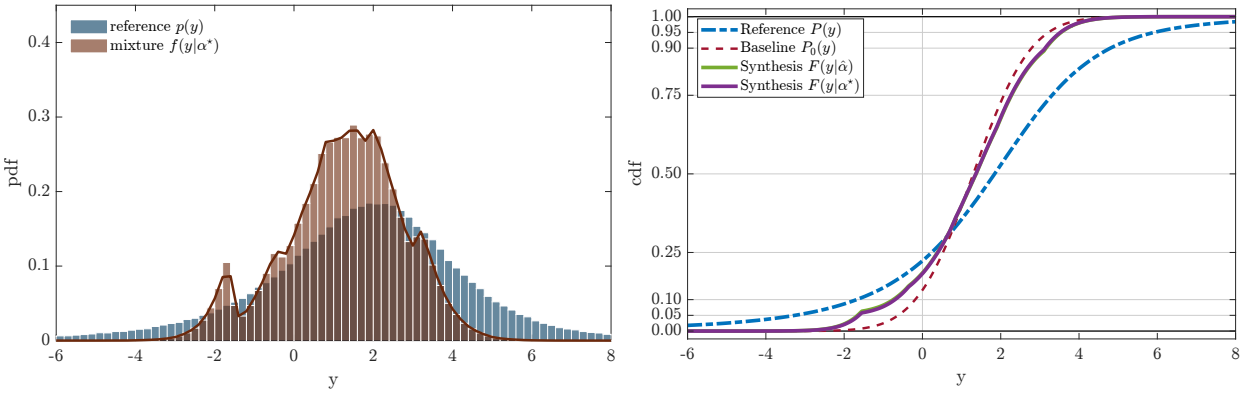
6.5 Scenario Synthesis based on P15, P50 and P85

As earlier noted, the methodology admits specification of multiple features of the scenarios so long as they can be represented as expectations under implicit scenario distributions. The case of multiple percentiles is practically key, and we visit this setting with summaries of further analysis in the TB 2007 context.

Figure 4 and Table 6 report the results when scenario p.d.f.s are tilted versions of the baseline that match the scenario-specific P15, P50, and P85. Here the scenario P15 and P85 are computed assuming distances of tail percentiles from median agree with the baseline: $\text{P15}_j = \text{P50}_j - (\text{P50}_0 - \text{P15}_0)$ and $\text{P85}_j = \text{P50}_j + (\text{P85}_0 - \text{P50}_0)$. The P15 and P85 for the scenarios in Table 6 are similar to those in Table 4 when the scenario is close to the baseline; they are naturally more discordant with increasing departure of the scenario percentiles from those of baseline (e.g., $\mathcal{S}_2$). Then, optimal weights and EMR result are quite similar to those in Table 4. Other specifications of the P15 and P85 uncertainties– specifications that are founded in scenario considerations– may be quite different than the synthetic choices here, of course, and can be expected to lead to different results. A main point is that the methodology is open to– and trivially applied to– scenario specifications in terms

of multiple percentiles, with negligible analytic and computational burden. Such specifications are increasingly common in application, and to be encouraged in policy research moving forward.

Figure 4: p.d.f. and c.d.f of scenario synthesis and NY Fed reference
Tilted scenario distributions obtained using three percentiles — December 2007 Tealbook



Left: p.d.f. of the reference and scenario synthesis. *Right:* c.d.f. of the reference, baseline and scenario synthesis, the latter comparing results using $\widehat{\boldsymbol{\alpha}}$ and $\boldsymbol{\alpha}^*$ for the synthesis; again, the latter are effectively indistinguishable.

Table 6: Synthesis based on P15, P50 and P85 information: TB 2007, NY Fed reference

j	Scenario $\mathcal{S}$	P15$_j$	P50$_j$	P85$_j$	ET$_j$%	IS$_j$%	π_j^*	$\widehat{\alpha}_j$	α_j^*
0	Baseline	0.1	1.3	2.5	100.0	62.6	0.41	0.30	0.26
1	Greater housing correction	−0.2	1.0	2.2	92.3	57.2	0.39	0.00	0.01
2	Credit crunch	−1.5	−0.3	0.9	20.0	31.6	0.32	0.10	0.11
3	Stronger domestic demand	0.5	1.7	2.9	90.1	65.9	0.42	0.00	0.07
4	Better export performance	0.7	1.9	3.1	79.3	66.1	0.42	0.30	0.26
5	Greater cost pressure	0.0	1.2	2.4	99.4	61.2	0.40	0.00	0.01
6	Market-based Fed Funds rate	0.4	1.6	2.8	96.0	65.0	0.42	0.00	0.03
7	backstop	−1.5	1.4	3.1	27.7	62.4	0.43	0.30	0.26
	$f(\mathbf{y}\|\widehat{\boldsymbol{\alpha}})$	−0.3	1.4	2.8		73.0	0.44		
	$f(\mathbf{y}\|\boldsymbol{\alpha}^*)$	−0.3	1.4	2.8		72.7	0.44		

Details as in Table 4.

7 Distributional Forecasts and Judgment

At a general level, our focus is on use and reconciliation of information from statistical models and judgmental sources. Analysis is directional in that the specific goals are to assess judgmentally derived scenarios against a statistical reference distribution. In the broader context, the reverse is also of interest; that is, investigation of how a statistical forecast distribution may be "tilted" in a direction deemed important from a judgmental point of view. The latter can often be proxy for information external to that underlying the statistical model.

A key context is that of unique, unexpected events and shifts in the structure of the economy that go well beyond existing model structure and assumptions. While structural economic models

may provide a formal basis for longer-term adaptation, fully modeling the implications of regime shifts on the structure of the economy takes time. Short-term, judgmental adjustments can be most valuable for real-time decision making. Indeed, judgment plays a dominant role in decision making in other areas, such as among investment professionals in macroeconomic trading. In monetary policy settings, quantitative macroeconomic models provide a firmer basis, but undesirable policy recommendations from models are often attributed to persistent forecast errors. A "good" policy maker would intervene to input judgment to address this.

Subjective information sources include surveys, market intelligence, and stress test outputs. Some comments on each are germane.

i. Surveys. Perhaps the key example is the US Survey of Professional Forecasters (SPF) a well-established source of community-wide forecast information. SPF now collates probabilistic forecasts on predefined bins for outcomes (Croushore, 1993; Del Negro et al., 2023). Similar regular surveys are conducted by ECB, the Bank of England, and other institutions.

ii. Market Intelligence. Large and detailed information sets are commonly collected by central banks in order to inform policy makers on many dimensions of economic and financial market developments outside the scope of well-adopted structural macroeconomic models. Such models, aiming to reduce complexity and lead to openness and interpretation in economic terms, inevitably lack the ability to reflect the full complexity of structural changes or nonlinear dynamics that become practically relevant in more unusual circumstances. More complex economic and financial market intelligence– in the form of summary external forecast information and judgment– can then add real value, if recognized and appropriately integrated with the model-based forecasts.

iii. Stress testing. Initially developed by the IMF in the 1990s to assess financial system resilience, stress testing was later broadly adopted to assess macroeconomic risks from banking distress following the Great Financial Crisis (Adrian et al., 2020). All major financial regulators now design and publish macroeconomic stress scenarios and assess financial stability relative to those scenarios. This typically includes scenarios of major downturns in macroeconomic aggregates such as real activity, inflation, and financial conditions. Scenario design emphasizes extreme economic and financial circumstances; the resulting outputs can provide a basis for judgmental modification of forecasts from the established reference econometric models that are not designed or customized to quickly and easily address such circumstances.

The overall question is that of intervention in the statistical model to incorporate such external information. The concept is long recognized and much methodology exists and is used in other areas of forecasting, such as commercial and financial applications (e.g., West and Harrison, 1986; West and Harrison, 1989; Black and Litterman, 1991; West and Harrison, 1997, chap. 11; West, 2024). However, emphasizing and formalizing the question with respect to policy applications is highlighted and of renewed interest here.

Beyond contextual connections with the main theme of the current paper, key technical features of our scenario synthesis methodology relate directly to these complementary interests. From Section 5 and generalizing the notation there, each of the constructed scenario p.d.f.s has the form $p_j(\mathbf{y}) \propto w_j(\mathbf{y})p(\mathbf{y})$ based on the statistical reference p.d.f. $p(\mathbf{y})$ and scenario-specific weight– or tilting– function $w_j(\mathbf{y})$. The latter is $w_j(\mathbf{y}) \propto w_0(\mathbf{y})\exp\{\boldsymbol{\tau}_j'\mathbf{s}_j(\mathbf{y})\}$ involving: (i) the ET term $\exp\{\boldsymbol{\tau}_j'\mathbf{s}_j(\mathbf{y})\}$ used to define $p_j(\cdot)$ using the baseline and partial scenario information; and (ii) the baseline-reference IS weight function $w_0(\mathbf{y}) \propto p_0(\mathbf{y})/p(\mathbf{y})$. The Monte Carlo methodology uses the discrete versions over the reference random samples $\mathbf{y}^i$, i.e., $w_j^i = w_j(\mathbf{y}^i)$ for $j = 0{:}J$.

The normalized scenario p.d.f.s are then $p_j(\mathbf{y}) = c_j w_j(\mathbf{y})p(\mathbf{y})$ where the c_j are just normalizing

constants. Thus, explicitly, the partial judgmental information that $\mathcal{S}_j$ encodes leads to a weighted modification of the statistical reference; on $\mathcal{S}_j$ alone, this can be regarded as the scenario-tilted version of the model-based forecast. This indicates that the overall question of conditioning a model-based forecast on what may be quite distinct forms of judgmental information is intimately addressed within our framework. It also follows that the BPS-justified scenario mixture p.d.f. is $f(\mathbf{y}|\boldsymbol{\alpha}) = w(\mathbf{y}|\boldsymbol{\alpha})p(\mathbf{y})$ where $w(\mathbf{y}|\boldsymbol{\alpha}) = \sum_{j=0:J} \alpha_j c_j w_j(\mathbf{y})$. This is true for any $\boldsymbol{\alpha}$, not just the EMR-optimized value central to our scenario synthesis goals. In other contexts, such as the above settings of modifying the reference $p(\mathbf{y})$ with judgment-based information summaries, this allows for context-specific specification of relative scenario weightings. Importantly, scenarios can address multiple aspects of the forecast distribution, including location shifts, scale and skewness perturbations– both within any one scenario and with diversity across a scenario set. This may be particularly important to extension and evaluation of this approach in areas such as stress testing.

8 Summary Comments

The formal assessment and integration of partial scenario information with statistical forecast distributions is of interest in a range of policymaking settings. Our approach has been motivated by the monetary policy process, where policy decisions are firmly rooted in macroeconomic forecasts that involve not only the baseline forecast, but also alternative risk scenarios. The methodology offers a concrete and straightforward approach to evaluating baseline and judgmental scenario assessments– with their intuitive and easily communicated bases– against more formal statistical density forecasts of risk. Applications to the monetary policy process are highlighted, and offer a new frontier for practical yet rigorous policy making.

We believe the methodology will have broad appeal in other applications. For example, the IMF continuously monitors global financial stability, and publishes a formal biannual GaR-based global financial stability assessment in its Global Financial Stability Report. The statistical approach can be compared directly– and in a quantitatively meaningful manner– with the scenario-based risk assessment of the IMF's World Economic Outlook, published on the same schedule as the GFSR.

The approach can also be readily applied to other areas of institutional risk management and contexts such as portfolio choice applications. In risk management, as in financial institution supervisory stress testing, both scenario-based approaches and more statistical approaches are commonly deployed. Our framework and methodology offers a novel way to evaluate and integrate those two avenues in a concrete fashion. In portfolio allocation decision making, the role of priors is fundamental and features commonly in allocation decisions, yet the bridge between intuitive scenario based approaches and statistical modeling of return forecasts has received little attention. Again our framework offers steps ahead in this regard. Beyond these areas, there are also opportunities in commercial revenue and supply chain forecasting where ranges of forms of external/subjective information are often assessed in the context of formal models with a view to eventual decisions.

Forms of scenario information that may feed-into new applications are information sets with more than a few candidate percentiles of forecast distributions under any assumed scenario. If a scenario is specified in terms of a larger number of percentiles, then analysis begins to approximate that given a fully-specified scenario p.d.f. $p_j(\mathbf{y})$. This is certainly of methodological interest, and may represent applied interests in settings where the scenarios are effectively replaced by forecast distributions from alternative/competing models. This latter setting is closer to the existing setting of BPS where multiple predictive distributions are considered by a Bayesian decision maker, and

define analogue information to condition an initial reference forecast distribution. Some of the technical developments here are rather different– and complementary– to the general setting of BPS, but open up new questions for potential development and exploitation in forecast model comparison and synthesis. There are also questions of extension of the technical approach to address synthesis based on other forms of scenario information, e.g., point forecasts that are regarded as subjectively assessed modes or means rather than medians, and uncertainty in scenario information, e.g., percentiles provided with some notational $\pm$ uncertainties. The general ET framework in principle applies to such contexts, though details are to be developed for exploitation in any new applied context in which such scenario information sets arise.

We have noted that the methodology is applicable with $\mathbf{y}$ in several dimensions. Elements of $\mathbf{y}$ can include multiple economic indicators (real growth, inflation, unemployment, etc.,) as well as– anchored at a current time period, such as the end of the current quarter– multiple time periods ahead, such as the coming eight quarters. Scenario information to define (uncertain) constraints on forecasts of the state of the macroeconomy over multiple future time periods can than generate a range of scenarios. Technically and computationally, the methodology here extends immediately. We have experience with such extensions, and recognize questions that arise due to increasing dimension of $\mathbf{y}$. In technical essentials, the main questions there are not new, but have to do with scalability of importance sampling methodology with dimension, and of its close technical ally entropic tilting with increasing dimension of the underlying Bayesian decision-analytic utility (a.k.a. score) functions. These questions are addressed in all applications of these general approaches, and will need to be addressed in context– in specific applied settings of scenario synthesis.

Acknowledgments

The authors thank colleagues at the Federal Reserve and the IMF– including Gianni Amisano, Matthias Paustian, Sheheryar Malik, Jason Wu, and Pierre-Olivier Gourinchas– for multiple discussions and feedback on the general area and a range of specific topics. We also thank Todd Clark of the Center for Financial Economics at Johns Hopkins University for his detailed, thoughtful comments on a first draft of our paper. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve System, the Federal Open Market Committee, or the International Monetary Fund, its Management, or its Executive Directors.

References

Adams, P. A., T. Adrian, N. Boyarchenko, and D. Giannone (2021). Forecasting macroeconomic risks. *International Journal of Forecasting 37*(3), 1173–1191.

Adrian, T. and N. Boyarchenko (2012). Intermediary leverage cycles and financial stability. Staff Report 567, Federal Reserve Bank of New York.

Adrian, T., N. Boyarchenko, and D. Giannone (2016). Vulnerable growth. Staff Report 794, Federal Reserve Bank of New York.

Adrian, T., N. Boyarchenko, and D. Giannone (2019). Vulnerable growth. *American Economic Review 109*(4), 1263–1289.

Adrian, T., N. Boyarchenko, and D. Giannone (2021). Multimodality in macrofinancial dynamics. *International Economic Review 62*(2), 861–886.

Adrian, T., F. Grinberg, N. Liang, S. Malik, and J. Yu (2022). The term structure of Growth-at-Risk. *American Economic Journal: Macroeconomics 14*(3), 283–323.

Adrian, T., J. Morsink, and L. B. Schumacher (2020). Stress testing at the IMF: A framework for macroprudential analysis. Departmental Paper 2020/016, International Monetary Fund.

Aikman, D., J. Bridges, S. H. Hoke, C. O'Neill, and A. Raja (2019). Credit, capital and crises: a GDP-at-Risk approach. Staff Working Paper 824, Bank of England.

Alessandri, P., L. D. Vecchio, and A. Miglietta (2019). Financial conditions and 'Growth at Risk' in Italy. Economic Working Paper 1242, Bank of Italy, Economic Research and International Relations Area.

Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society (Ser. B) 36*(1), 99–102.

Anesti, N., M. Garofalo, S. Lloyd, E. Manuel, and J. Reynolds (2023). Unknown measures: Assessing uncertainty around UK inflation using a new Inflation-at-Risk model. Bank Underground, Bank of England.

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

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society (Ser. B) 65*(2), 367–389.

Azzalini, A. and A. Capitanio (2013). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press.

Bernanke, B. (2023). Forecasting for monetary policy making and communication at the Bank of England: A review. Report to the BoE Independent Evaluation Office, Bank of England.

Black, F. and R. B. Litterman (1991). Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income 1*(2), 7–18.

Boyarchenko, N., R. K. Crump, L. Elias, and I. L. Gaffney (2023). Look out for Outlook-at-Risk. Liberty Street Economics 20230517, Federal Reserve Bank of New York.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization (additional exercises)*. Cambridge University Press.

Britton, E., P. Fisher, and J. Whitley (1998). The inflation report projections: Understanding the fan chart. Quarterly Bulletin Q1, Bank of England.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences 106*(30), 12267–12272.

Brunnermeier, M. K. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104*(2), 379–421.

Caldara, D., D. Cascaldi-Garcia, P. Cuba-Borda, and F. Loria (2021). Understanding Growth-at-Risk: A Markov switching approach. *SSRN Electronic Journal*. doi:10.2139/ssrn.3992793.

Carriero, A., T. E. Clark, and M. Marcellino (2024). Capturing macro-economic tail risks with Bayesian vector autoregressions. *Journal of Money, Credit and Banking 56*(5), 1099–1127.

Chernis, T., G. Koop, E. Tallman, and M. West (2024). Decision synthesis in monetary policy. Bank of Canada, Staff Working Paper 2024-30. arXiv:2406.03321.

Clark, T. E., G. Ganics, and E. Mertens (2022). What is the predictive value of SPF point and density forecasts? Working paper no. 22-37, Federal Reserve Bank of Cleveland. doi:10.26509/frbc-wp-202237.

Conflitti, C., C. De Mol, and D. Giannone (2015). Optimal combination of survey forecasts. *International Journal of Forecasting 31*(4), 1096–1103.

Croushore, D. (1993). Introducing: The survey of professional forecasters. Business Review 3/1993, Federal Reserve Bank of Philadelphia.

Crump, R. K., S. Eusepi, D. Giannone, E. Qian, and A. M. Sbordone (2025). A large Bayesian VAR of the United States economy. *International Journal of Central Banking -*, –.

Crump, R. K., M. Everaert, D. Giannone, and C. S. Hundtofte (2024). Changing risk-return profiles. In M. Barigozzi, S. Hörmann, and D. Paindaveine (Eds.), *Recent Advances in Econometrics and Statistics: Festschrift in Honour of Marc Hallin*, pp. 283–302. Springer.

De Mol, C. (2024). Multiplicative algorithms for density combination and deconvolution. In *Recent Advances in Econometrics and Statistics: Festschrift in Honour of Marc Hallin*, pp. 493–510. Springer.

Del Negro, M., F. Bassetti, and R. Casarin (2023). Inference on probabilistic surveys in macroeconomics with an application to the evolution of uncertainty in the Survey of Professional Forecasters during the COVID pandemic. In W. van der Klaauw, G. Topa, and R. Bachmann (Eds.), *Handbook of Economic Expectations*. Elsevier.

Diebold, F. X., M. Shin, and B. Zhang (2023). On the aggregation of probability assessments: Regularized mixtures of predictive densities for Eurozone inflation and real interest rates. *Journal of Econometrics 237*(2), 105321.

Eguren-Martin, F., S. Kösem, G. Maia, and A. Sokol (2024). Targeted financial conditions indices and Growth-at-Risk. Staff Working Paper 1084, Bank of England.

Engstrom, E. and M. Gonzalez-Astudillo (2017). Time variation in upside and downside risks to the staff baseline forecast. Staff Memo to the Federal Open Market Committee, Board of Governors of the Federal Reserve System.

Federal Reserve Board (2007). Report to the FOMC on Economic Conditions and Monetary Policy. Part 1– Current Economic and Financial Conditions: Summary and Outlook. December 5, 2007, Board of Governors of the Federal Reserve System.

Federal Reserve Board (2018). Report to the FOMC on Economic Conditions and Monetary Policy. Book A–Economic and Financial Conditions: Outlook, Risks, and Policy Strategies. December 7, 2018, Board of Governors of the Federal Reserve System.

Fernández-Villaverde, J., P. Guerrón-Quintana, J. F. Rubio-Ramírez, and M. Uribe (2011). Risk matters: The real effects of volatility shocks. *American Economic Review 101*(6), 2530–2561.

Fernández-Villaverde, J., S. Hurtado, and G. Nuno (2023). Financial frictions and the wealth distribution. *Econometrica 91*(3), 869–901.

Fernández-Villaverde, J., F. Mandelman, Y. Yu, and F. Zanetti (2024). Search complementarities, aggregate fluctuations, and fiscal policy. *Review of Economic Studies*. rdae053.

Figueres, J. M. and M. Jarociński (2020). Vulnerable growth in the Euro area: Measuring the financial conditions. *Economics Letters 191*(C), 109–126.

Giannone, D., M. Lenza, and G. E. Primiceri (2021). Economic predictions with big data: The illusion of sparsity. *Econometrica 89*(5), 2409–2437.

Gruber, L. F. and M. West (2016). GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models. *Bayesian Analysis 11*(1), 125–149.

Gruber, L. F. and M. West (2017). Bayesian forecasting and scalable multivariate volatility analysis using simultaneous graphical dynamic linear models. *Econometrics and Statistics 3*(C), 3–22.

Hafemann, L. (2023). House prices at risk: A framework for assessing vulnerabilities in the German housing market. Technical Paper 7/2023, Deutsche Bundesbank.

He, Z. and A. Krishnamurthy (2012). A model of capital and crises. *Review of Economic Studies 79*(2), 735–777.

IMF (2017). *Global financial stability report: Is growth a risk?* International Monetary Fund.

Johnson, M. C. and M. West (2025). Bayesian predictive synthesis with outcome-dependent pools. *Statistical Science 40*(1), 109–127.

Jondeau, E., P. Poncet, and C. Rebillard (2022). Are financial variables useful to complement GDP nowcasting? Eco Notepad, Banque de France.

Justiniano, A. and G. E. Primiceri (2008). The time-varying volatility of macroeconomic fluctuations. *American Economic Review 98*(3), 604–641.

Kiley, M. T. (2022). Unemployment risk. *Journal of Money, Credit and Banking 54*(5), 1407–1424.

Koop, G., S. McIntyre, and J. Mitchell (2019). UK regional nowcasting using a mixed frequency vector autoregressive model with entropic tilting. *Journal of the Royal Statistical Society (Ser. A) 183*(1), 91–119.

Krüger, F., T. E. Clark, and F. Ravazzolo (2017). Using entropic tilting to combine BVAR forecasts with external nowcasts. *Journal of Business and Economic Statistics 35*(3), 470–485.

Leeper, E. M. and T. Zha (2003). Modest policy interventions. *Journal of Monetary Economicss 50*(8), 1673–1700.

Lenza, M., I. Moutachaker, and J. Paredes (2023). Density forecasts of inflation: A quantile regression forest approach. Working Paper Series 2830, European Central Bank.

Lin, L., C. Chan, and M. West (2016). Discriminative variable subsets in Bayesian classification with mixture models, with application in flow cytometry studies. *Biostatistics 17*(1), 40–53.

López-Salido, D. and F. Loria (2024). Inflation at risk. *Journal of Monetary Economics 145*(S), 103570.

Matlab (2024). Matlab Optimization Toolbox (version 24.1, r2024a). https://www.mathworks.com.

McAlinn, K. and M. West (2019). Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics 210*(1), 155–169.

Metaxoglou, K., D. Pettenuzzo, and A. Smith (2018). Option-implied equity premium predictions via entropic tilting. *Journal of Financial Econometrics 17*(4), 559–586.

Plagborg-Moller, M., L. Reichlin, G. Ricco, and T. Hasenzagl (2020). When is growth at risk? *Brookings Papers on Economic Activity 51*(1), 167–229.

Pujadas, A. M., L. Hospido, and J. M. Montero (2022). House prices at risk: An empirical approach to downside risks in the Spanish housing market. Working Paper 2244, Banco de España.

Robertson, J. C., E. W. Tallman, and C. H. Whiteman (2005). Forecasting using relative entropy. *Journal of Money, Credit, and Banking 37*(3), 383–401.

Tallman, E. and M. West (2022). On entropic tilting and predictive conditioning. *Supporting material for* Tallman and West (2023). arxiv:2207.10013.

Tallman, E. and M. West (2023). Bayesian predictive decision synthesis. *Journal of the Royal Statistical Society (Ser. B) 86*(2), 340–363.

Tallman, E. and M. West (2025). Predictive decision synthesis for portfolios: Betting on better models. In S. Mazur and P. Österhol (Eds.), *Recent Developments in Bayesian Econometrics and Their Applications*. Springer. arXiv:2405.01598.

West, M. (1987). On scale mixtures of normal distributions. *Biometrika 74*(3), 646–648.

West, M. (2024). Perspectives on constrained forecasting. *Bayesian Analysis 19*(4), 1013–1039.

West, M. and P. J. Harrison (1986). Monitoring and adaptation in Bayesian forecasting models. *Journal of the American Statistical Association 81*(395), 741–750.

West, M. and P. J. Harrison (1989). Subjective intervention in formal models. *Journal of Forecasting 8*(1), 33–53.

West, M. and P. J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2nd ed.). Springer.

Appendix A Relating KL Divergences to EMR

A.1 Lower Bounds on EMR

As noted in Section 4.2, the bound $\pi_{pf} \geq 1/[1 + \exp\{\mathrm{KL}(p\|f)\}]$ has empirical support in specific examples. This is, of course, not true in general, though seems suggested under certain, practically relevant conditions on $y \sim p(\mathbf{y})$. The implied distribution of $k(\mathbf{y}) = \log\{p(\mathbf{y})/f(\mathbf{y})\}$ has mean $\mathrm{E}[k(\mathbf{y})|\mathcal{H}_p] = \mathrm{KL}(p\|f) \geq 0$ with equality only when $k(\mathbf{y}) = 0$ for all $\mathbf{y}$. The bound is conjectured to hold when the distribution of $k(\mathbf{y})$ has finite, positive mean, is unimodal with $\Pr[k(\mathbf{y}) > 0] > 0.5$, and has p.d.f. tail decay on $k(\mathbf{y}) < 0$ no heavier than that on $k(\mathbf{y}) > 0$. An exact proof for more restricted cases is available, as follows.

Simplifying notation, the real-valued quantity k replaces $k(\mathbf{y})$. The focus is on $\pi_{pf} = \mathrm{E}_k[\pi(k)]$ where $\pi(k) = 1/\{1+\exp(k)\}$ and k has some p.d.f. $g(k)$ with finite mean $m > 0$. Here $\mathrm{E}_k[\cdot]$ denotes expectation with respect to $k \sim g(k)$. The following theory draws on standard results concerning scale mixtures of normals (Andrews and Mallows, 1974; West, 1987).

Suppose that $g(k)$ is continuous, symmetric and unimodal with mode and finite mean m. Then $g(k)$ is a normal scale mixture: $g(k) = \mathrm{E}_v[v^{-1}\phi\{v^{-1}(k-m)\}]$ where $\phi(\cdot)$ is the standard normal p.d.f., v is a random scale parameter and $\mathrm{E}_v[\cdot]$ denotes expectation with respect to its distribution.

Then, recognize that $\pi(k) = 1/\{1 + \exp(k)\}$ is the survival function of the standard univariate logistic distribution for real-valued k. The logistic distribution is also a normal scale mixture, so $\pi(k) = 1 - \mathrm{E}_u[\Phi(u^{-1}k)]$ where $\Phi(\cdot)$ is the standard normal c.d.f., u is the random scale parameter and $\mathrm{E}_u[\cdot]$ denotes expectation with respect to its distribution. Thus $\pi(m) = 1 - \mathrm{E}_u[\Phi(u^{-1}m)]$.

The above theory of the normal scale mixture structure $g(k)$ and $\pi(k)$ further implies that $\pi_{pf} = 1 - \mathrm{E}_{v,u}[\int_k \Phi(u^{-1}k)v^{-1}\phi\{v^{-1}(k-m)\}dk]$ with expectation over v, u (in which, implicitly, $u \perp\!\!\!\perp v$). Routine normal theory yields $\pi_{pf} = 1 - \mathrm{E}_{v,u}[\Phi(w^{-1}m)]$ where $w = \sqrt{v^2 + u^2}$. Hence $\mathrm{E}_k[\pi(k)] - \pi(m) = \mathrm{E}_{v,u}[\Phi(u^{-1}m) - \Phi(w^{-1}m)]$. Now, $m \geq 0$ and $w > u$ so that $\Phi(u^{-1}m) - \Phi(w^{-1}m) \geq 0$ implying that $\mathrm{E}_k[\pi(k)] \geq \pi(m)$, as required. This inequality is strict unless $k = v = 0$.

The analysis above may extend to more general cases when $g(k)$ is not symmetric. Suppose, for example, that $g(k)$ is a scale mixture of skew-normal distributions (Azzalini and Capitanio, 2013). This is a rich class of unimodal distributions with ranges of asymmetries; it includes the above symmetric distributions as special cases. Convolutions of skew-normals with normals are skew-normals, so it is reasonable to ask if the above development generalizes. There may be broader generalization so long as $g(k)$ is unimodal with $m > 0$ and/or $\Pr(k \geq 0) > 0.5$. This is an aside and beyond current scope, but suggests further theoretical study. Importantly, the above discussion does not extend at all to cases– including many practical cases– when the expectation of k (defining the KL divergence) does not exist and when the distribution of k is less regular and even multimodal.

As another aside note, this also provides interpretation of KL on the probabilistic concordance scale in cases when the bound is assured; in such cases, $\kappa_{pf}, \kappa_{fp} \leq \log\{(1 - \pi_{pf})/\pi_{pf}\}$.

A.2 Approximation and Bounds

The link of EMR to KL is further illuminated in the 1st-order Taylor series approximation of the function $1/\{1 + \exp(k)\}$ at $k = 0$, namely $1/\{1 + \exp(k)\} \approx (2 - k)/4$. This is an exact lower bound on $1/\{1 + \exp(k)\}$ for $k \geq 0$ and an exact upper bound for $k \leq 0$. See this as follows. First, $\pi(k) - (2 - k)/4$ is positive on $k > 0$ if, and only if, $g(k) > 0$ where $g(k) = k + 2 - (2 - k)\exp(k)$. Calculus shows that $g(k)$ is strictly increasing for all k, and, of course, $g(0) = 0$, hence the lower

bound arises for $k > 0$. Second, $\pi(k) - (2-k)/4$ is negative on $k < 0$ if, and only if, $g(k) < 0$, so the upper bound is implied on that range. This approximation is very accurate over $|k| \leq 0.5$ where $1/\{1+\exp(k)\} \geq 0.38$, i.e., in cases of relatively close concordance; the absolute error is less than 0.68% on $|k| \leq 0.5$. Hence, if $\mathbf{y} \sim p(\mathbf{y})$ and the implied distribution of $k(\mathbf{y})$ heavily favors such ranges, then $\pi_{pf} \approx \{2-\kappa_{pf}\}/4$. On this basis, choosing $f(\cdot)$ to maximize π_{pf} is again approximately the (symmetrized) KL divergence minimizing solution. In the following example highlights values of $\pi_{pf} \geq 0.4$ as practically relevant, as lower values are strong indications of lack of concordance.

A.3 A Simple Example

A simple example relates the range of π_{pf} to the familiar, interpretable measure of expected sample size (ESS) from Monte Carlo (MC) analysis using importance sampling (IS), in addition to KL. Take $\mathbf{y} = y$ to be scalar with $p(y) = \text{N}(0,1)$, standard normal, and $f(y) = \text{N}(a,1)$ for some mean $a \geq 0$. IS with proposal $p(y)$ and target $f(y)$ as target leads to MC integration based on the resulting weighted average approximations. A random sample $y_i \sim p(y)$, $(i = 1{:}n)$, leads to IS weights $w_i \propto f(y_i)/p(y_i)$ subject to summing to 1. The resulting MC approximation to EMR is $\pi_{pf} \approx \sum_{i=1:n} w_i/(1+nw_i)$. The IS effective sample size as a percentage is $\text{ESS} = n^{-1}100/\sum_{i=1:n} w_i^2$. Also, in this simple example $\text{KL}(p\|f) = \text{KL}(f\|p) = \kappa_{pf} = a^2/2$.

Figure 5 comes from an example with $n = 10^6$ and across a range of values of $a > 0$. For $a \leq 1$, $\text{ESS} \geq 40\%, \pi_{pf} \geq 0.40$ and is roughly linear in ESS up to its maximum of 0.5. For practical purposes and extrapolating from this interpretable example– also supported by other empirical examples– $\pi_{pf} \geq 0.4$ or so is expected unless $f(\cdot)$ and $p(\cdot)$ are quite substantially discordant. From Section 4.2, $\pi_{pf} \geq 0.40$ links to the accurate approximation $\pi_{pf} \approx 1/\{1+\exp(\kappa_{pf})\}$ with $\kappa_{pf} \leq 0.4$.

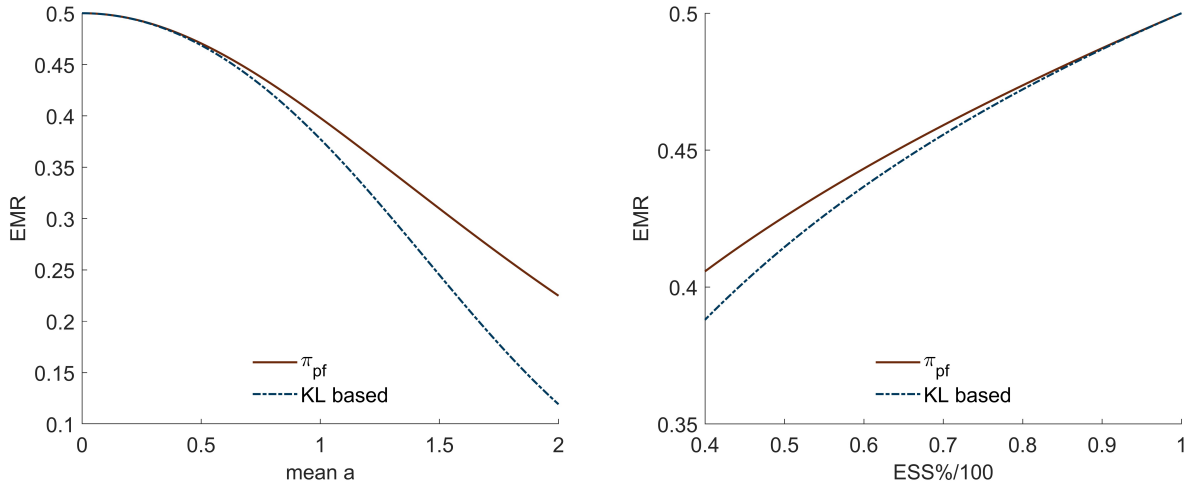


Figure 5: Predictive Concordance Example: EMR, ESS and KL-based lower bound when $p(y) = \text{N}(0,1)$ and $f(y) = \text{N}(a,1)$ for a range of values of a.

Appendix B Prior/Regularization Parameter Specification

We have prior $\boldsymbol{\alpha} \sim \text{Dir}(\mathbf{1}(1+\epsilon))$ and aim to calibrate the choice of small ϵ. We discuss this by analogy with canonical setting for Dirichlet prior/posterior distributions, i.e., multinomial sampling. With

$\boldsymbol{\alpha}$ representing scenario probabilities, the least informative multinomial sample is just one draw from one of the scenarios; the implied Dirichlet posterior then has parameter updated by $+1$ in one element only. With no loss of generality, suppose a single outcome is known to come from the baseline; this single draw posterior is then Dir$(\mathbf{1}(1+\epsilon)+\mathrm{e})$ where $\mathrm{e} = (1, 0, \ldots, 0)'$. Now, to reflect a minimally informative setting, suppose the posterior is modified to Dir$(\mathbf{1}(1+\epsilon)+x\mathrm{e})$ for some very small, positive x. This can be regarded as the posterior under an imaginary fractional observation; for example, $x = 0.01$ says the information content of the posterior relative to the prior is 1% of that arising on observing a single multinomial draw.

Under this posterior with specified x, the prior mode $1/(J+1)$ increases to posterior mode $\alpha_0^* = (\epsilon + x)/\{(J+1)\epsilon + x\}$ on $\mathcal{S}_0$, and decreases to $\alpha_j^* = \epsilon/\{(J+1)\epsilon + x\}$ on the other scenarios $j > 0$. In this minimal information context it is rationale to limit this latter "shrinkage towards zero" and we reflect this by asking that $\alpha_j^* \geq p/(J+1)$ for some fractional reduction $p \in (0, 1)$. This implies $\epsilon \geq cx/(J+1)$ where $c = p/(1-p)$. Here c is explicitly a lower bound on the reduction from prior to posterior *odds* on $\mathcal{S}_j$ for $j > 0$ given the minimal information of a single outcome under $\mathcal{S}_0$. For example, the choice $c = 0.5$ limits this odds reduction to no more than 50%. The choices $x = 0.01$ and $c = 0.5$ imply $\epsilon \geq 0.005/(J+1)$, and this value is recommended as a default.

Appendix C Optimization of Scenario Mixtures

For any $\mathbf{y}$ define the $(J+1)-$vector $\mathbf{p}(\mathbf{y}) = [p_0(\mathbf{y}), \ldots, p_J(\mathbf{y})]'$. It is then easily shown that derivatives of EMR in eqn. (4) are

$$\mathbf{h}(\boldsymbol{\alpha}) \equiv \frac{\delta \pi_{pf}(\boldsymbol{\alpha})}{\delta \boldsymbol{\alpha}} = \int_{\mathbf{y}} \mathbf{p}(\mathbf{y}) h(\mathbf{y}|\boldsymbol{\alpha}) d\mathbf{y} \quad \text{and} \quad \mathbf{H}(\boldsymbol{\alpha}) \equiv \frac{\delta^2 \pi_{pf}(\boldsymbol{\alpha})}{\delta \boldsymbol{\alpha} \boldsymbol{\alpha}'} = -2 \int_{\mathbf{y}} \mathbf{p}(\mathbf{y}) \mathbf{p}(\mathbf{y})' H(\mathbf{y}|\boldsymbol{\alpha}) d\mathbf{y}$$

where $h(\mathbf{y}|\boldsymbol{\alpha}) = p(\mathbf{y})^2/\{p(\mathbf{y}) + f(\mathbf{y}|\boldsymbol{\alpha})\}^2$ and $H(\mathbf{y}|\boldsymbol{\alpha}) = h(\mathbf{y}|\boldsymbol{\alpha})/\{p(\mathbf{y}) + f(\mathbf{y}|\boldsymbol{\alpha})\}$. At any $\boldsymbol{\alpha}$ the Hessian matrix is $\mathbf{H}(\boldsymbol{\alpha}) = -2\,\mathrm{E}[\mathbf{p}(\mathbf{y})\mathbf{p}(\mathbf{y})' a(\mathbf{y}|\boldsymbol{\alpha})]$ where $a(\mathbf{y}|\boldsymbol{\alpha}) = p(\mathbf{y})/\{p(\mathbf{y}) + f(\mathbf{y}|\boldsymbol{\alpha})\}^3$ and the expectation is with respect to $\mathbf{y} \sim p(\cdot)$. Since $a(\mathbf{y}|\boldsymbol{\alpha}) > 0$ for all $\mathbf{y}, \boldsymbol{\alpha}$ the expectation is a positively weighted average of rank-one matrices $\mathbf{p}(\mathbf{y})\mathbf{p}(\mathbf{y})'$. Whether $\mathbf{y}$ is continuous or discrete (the latter with at least $J+2$ support points) $\mathbf{H}(\boldsymbol{\alpha})$ is full rank and strictly negative definite for all $\boldsymbol{\alpha}$.

It follows that maximizing $\pi_{pf}(\boldsymbol{\alpha})$ over the simple is a convex optimization problem with a unique maximizing value $\widehat{\boldsymbol{\alpha}}$; standard constrained optimization algorithms then apply. Further, the modification to add the prior penalty and define the log posterior objective function in eqn. (5) maintains convexity and ensures a unique posterior mode $\boldsymbol{\alpha}^*$ given any specified value of ϵ. Then, standard constrained, non-linear optimization methods (e.g., the default interior-point algorithm in the `fmincon` function in Matlab, 2024) apply and are fast and efficient.

As an aside but of some broader interest, the same approach shows that minimizing KL$(p\|f)$ or KL$(f\|p)$ (when finite) with respect to $\boldsymbol{\alpha}$ are also convex optimizations with unique solutions and similar characteristics. This also applies with any value of ϵ in the fully Bayesian version that adds the prior penalty based on a very diffuse but proper Dirichlet prior that supports non-zero α_j with probability one. This links to an existing literature on sparsity and stability of KL-optimal mixtures in the context forecast combination (e.g. Conflitti et al., 2015; Diebold et al., 2023; Crump et al., 2024; De Mol, 2024). Then, relative to EMR, the KL analysis typically leads to more zeros among the optimizing values of $\boldsymbol{\alpha}$ i.e., a sparser mixture more aggressively favoring just one or a small number of scenarios. This arises since KL involves expectations of the *unbounded* function

$\log\{p(\mathbf{y})/f(\mathbf{y}|\boldsymbol{\alpha})\}$ and is very dependent on behaviour of the tails of the two p.d.fs. This also relates to the caveat that, as noted in Section 4.2, KL divergence may simply be undefined in important practical contexts depending on the relative tail behavior of $p(\mathbf{y})$ and $f(\mathbf{y}|\boldsymbol{\alpha})$.

In contrast, EMR is more conservative (and numerically more robust) in discounting scenarios that are less concordant with the reference though not extremely so; this arises as EMR is based on expectations of the *bounded* function $1/\{1+p(\mathbf{y})/f(\mathbf{y}|\boldsymbol{\alpha})\}$ in eqn. (4). That said, the full shrinkage to boundaries of the simplex still arises and requires modest regularization as provided by the penalty induced under the minimally informative prior in the foundational Bayesian setting of Section 5.2.

Appendix D Summary of Computational Flow

1. Generate a large random sample $\mathbf{y}^i$, $(i = 1{:}n)$, from the reference $p(\mathbf{y})$, to define an importance sample for Monte Carlo evaluation of the baseline $p_0(\mathbf{y})$ and the scenarios $p_j(\mathbf{y})$.

2. Evaluate baseline IS weights $w_0^i \propto p_0(\mathbf{y}^i)/p(\mathbf{y}^i)$, subject to normalization.

3. Evaluate the scenario p.d.f.s $p_j(\mathbf{y})$ for $j > 0$.

   (a) If the scenario p.d.f.s are completely specified and can be evaluated, this is as in Step 2 above now applied to scenario p.d.f.s $p_j(\mathbf{y})$ instead of the baseline $p_0(\mathbf{y})$. For each $\mathcal{S}_j$ this delivers normalized IS weights w_j^i on the reference sample values.

   (b) If the scenarios are only partially specified, proceed as follows.

      i. Compute the scenario distributions using a random sample $\mathbf{x}^i$, $(i = 1{:}n)$, drawn from the baseline $p_0(\mathbf{x})$. Use this for MC evaluations of the integrals required to deliver tilting parameters for each $\mathcal{S}_j$ to minimally distort the baseline to match specified scenario medians, percentiles, etc.

      ii. Compute the implied tilting weights u_j^i on each of the reference sampled values $\mathbf{y}^i$.

      iii. Compute the implied ET-IS weights for each $p_j(\mathbf{y}^i)$ at the reference random sample values $\mathbf{y}^i$, namely the normalized weights $w_j^i \propto w_0^i u_j^i$, $(i = 1{:}n)$, for each $\mathcal{S}_j$.

4. Compute synthesis weights by optimizing eqn. (5) over $\boldsymbol{\alpha}$; this is modified with constraint $\alpha_0 \geq \max_{j=1:J} \alpha_j$ if the baseline is required to be the modal scenario as used in our examples. At each value of $\boldsymbol{\alpha}$ in iterations of a numerical optimization routine, direct MC integration evaluates EMR in eqn. (4). This is just the average over $i = 1{:}n$ of sampled EMR values $w_f^i(\boldsymbol{\alpha})/\{w_f^i(\boldsymbol{\alpha}) + w_p^i\}$ with implied scenario synthesis IS weights $w_f^i(\boldsymbol{\alpha}) = \sum_{j=0:J} \alpha_j w_j^i$ and uniform reference weights $w_p^i = 1/n$ for all $i = 1{:}n$.