FEDERAL RESERVE BANK *of* ATLANTA **WORKING PAPER SERIES**

On Model Aggregation and Forecast Combination

Nikolay Gospodinov and Esfandiar Maasoumi

Working Paper 2025-12
October 2025

Abstract: Policy makers express their views and decisions via the lens of a particular model or theory. But since any model is a highly stylized representation of the unknowable object of interest, all these models are inherently misspecified, and the resulting ambiguity injects uncertainty in the decision-making process. We argue that entropy-based aggregation is a convenient device to confront this uncertainty and summarize relevant information from a set of candidate models and forecasts. The proposed aggregation tends to robustify the decision-making process to various sources of risks and uncertainty. We find compelling evidence for the advantages of entropy-based aggregation for forecasting inflation.

JEL classification: C52, C53, E37, G11

Key words: model uncertainty, model aggregation, forecast combination, robust policy

https://doi.org/10.29338/wp2025-12

Nikolay Gospodinov (nikolay.gospodinov@atl.frb.org) is in the Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree Street NE, Atlanta, GA 30309-4470. Esfandiar Maasoumi (esfandiar.maasoumi@emory.edu) is in the Emory University Department of Economics, Rich Memorial Building 324, 1602 Fishburne Drive, Atlanta, GA 30322-2240.

1 Introduction

Competing theories of inflation are embodied in different models and opinions. If one theory is "true", all others are false. This means that all models are logically false in an empirical, inductive inference setting. How should a policy maker hedge against this endemic model uncertainty? Model aggregation is both informally and formally known to provide optimal hedging in a replicable and transparent manner. In contrast, the commonly employed methods of "model selection" are logically flawed as they seek to find "the" correct model.

Given that inflation and its reliable prediction are amongst the most pressing current policy questions, we propose and demonstrate that judicious aggregation of individual inflation forecasts produces dramatically superior performance compared to its component models of inflation, according to a wide set of risk criteria. This can provide a formal and transparent basis for policy maker decisions and pronouncements. Such "aggregation" is informally undertaken by policy maker faced with competing advice and input from a bewildering array of expert opinion and models. Our formal aggregation strategy points the way to an optimal approach to this informal undertaking, and is dominant over known alternatives. It also provides a broad view of model averaging and aggregation which identifies alternative formal and informal methods. As we show, our approach is easily implemented, and it can also accommodate machine learning models of inflation. Indeed, our method may be seen as not so "Artificial" Intelligence (AI) decision making when faced with competing analyses and forecasts.

We propose to aggregate point forecasts of a desired object, such as inflation, although the proposed method can be readily adapted for aggregation of density forecasts. Information theory offers powerful measures for contrasting distributions, and we emphasize that other metrics, such as the Bregman class of measures are transformations of the same information-theoretic measures. "Optimality" of our aggregators derives from these metrics that are optimized to find the best aggregators. This also reveals the subjectivity inherent in the choice of any optimization measure, and helps to identify what aspects of forecasts are being emphasized or which aspects of forecast errors are penalized. As is clear from our analysis, inflation is but an important example of wide applicability of the proposed techniques. We illustrate the advantages of entropy-based aggregation for forecasting U.S. core inflation.

The optimal aggregators also identify the statistically justified weights that are allocated

to each component model or opinion. This helps to reveal which theories are closer to the best empirically supported outcome. We discuss the applicability of optimal aggregation to a wide range of problems, including portfolio construction and asset pricing. Since our method is time adaptive, the weights change over time, revealing the relative efficacy of competing models at different economic stages and/or different policy regimes. We also argue that our optimization procedure offers a scientific basis for the informal way policy makers refer to diversity of opinions and models available to them.

2 Conceptual Framework

It is probably not an exaggeration to say that all financial and economic models are inherently misspecified as they are constructed to approximate a complex reality. This is often done intentionally as parsimonious models draw only partial or incomplete maps of the latent objects of interest either to emphasize particular aspects or because the underlying structure is completely unknown. As a result, it seems desirable to explicitly acknowledge the model uncertainty surrounding all investment, asset allocation, and policy decisions. The information-theoretic or maximum-entropy approach adapts naturally to the underlying model uncertainty and provides a consistent framework for aggregating information from different, partially specified models.

In this section, we review the analytical framework of Gospodinov and Maasoumi (2021) for robust aggregation based on information (entropy) theory. This framework capitalizes on insights from Maasoumi (1986) who proposed entropy-based aggregators that are constructed with the size distributions multiple indicators of a latest object, well-being. Furthermore, an axiomatic approach, based on postulating a few minimal properties (see, Maasoumi, 1993; Kobus, Kapera and Maasoumi, 2024) can provide a social decision-theoretic basis for policy maker's optimal aggregator, see Gospodinov and Maasoumi (2021) (see also Mudekereza, 2025). Finally, Gospodinov and Maasoumi (2025) use this approach to construct "ethical" measures of inflation which subjectively account for social and policy preferences that reflect the heterogeneous exposures of households at different income quintiles to various price components.

2.1 Divergence Measures

To introduce the maximum entropy principle, let P and Q be two probability measures with densities p and q with respect to a dominating measure μ; for example, two probability measures associated with two asset returns or the physical and risk-neutral probability measures. One way to measure the divergence between the two measures (Csiszár, 1972) is to solve the following optimization problem

$$D_\phi(P,Q) = \int \phi\left(\frac{dP}{dQ}\right) dQ,$$

where $\phi : \mathbb{R} \to [0, +\infty)$ is a convex, continuously differentiable function. The measure D_ϕ is nonnegative and $D_\phi(P,Q) = 0$ if and only if the two measures coincide, $P = Q$. For a choice of $\phi(\cdot)$, we use the Cressie-Read (Cressie and Read, 1984) power divergence family of functions

$$\phi(a) = \frac{a^{\rho+1} - 1}{\rho(\rho+1)} \text{ for } a \geq 0.$$

Different members of this family can be obtained for different values of the parameter ρ.[1] One celebrated member of this divergence family is the Kullback-Leibler divergence which is given by

$$\mathcal{KL}(P,Q) = \int \ln\left(\frac{p}{q}\right) q d\nu.$$

Our preferred divergence measure is the (scaled) Hellinger distance measure which is obtained by setting $\rho = -1/2$ or

$$\mathcal{H}(P,Q) = \frac{1}{2}\int \left(p^{1/2} - q^{1/2}\right)^2 d\nu.$$

2.2 General Aggregation

To introduce the main ideas, suppose that the policy makers observe an information signal about the underlying – possibly unknowable – object of interest $f(\cdot)$. These decision makers are endowed with models that represent their 'views' and are used as a convenient device to interpret and summarize an incoming information signal. More formally, suppose that there exists a finite dictionary $\mathcal{F} = \{f_1, ..., f_M\}$ of M candidate models or functions that approximate certain features of $f(\cdot)$. The goal is to construct, given the loss or risk function

[1] When this generalized entropy family is used to as an inequality measure (Maasoumi, 1986), the parameter $-\rho$ represents the degree of relative inequality aversion.

of the policy maker, a risk-minimized aggregator – a weighted sum of all candidate models – that generates the best approximate mapping between the aggregate and the individual candidate models, and reflects the uncertainty about $f(\cdot)$. Since we do not assume that the dictionary contains a "true" model, we refer to this aggregator as a pseudo-true aggregator that adapts to the least misspecified model in the dictionary.

Consider the flat simplex for a set of weights $w = (w_1, ..., w_M)$:

$$\mathcal{W}^M = \left\{ w \in \mathbb{R}^M : w_i \geq 0, \sum\nolimits_{i=1}^{M} w_i = 1 \right\}.$$

For a given risk function $\mathcal{R} : \mathcal{F} \rightarrow \mathbb{R}$, the pseudo-true aggregator of the candidates $\{f_1, ..., f_M\}$ is defined as

$$f_w^* = \underset{w \in \mathcal{W}^M}{\text{argmin}} \mathcal{R}(f^{(w)}, f).$$

The sample aggregator, denoted by $\tilde{f}^{(w)}$, is constructed by mimicking the pseudo-true aggregator using the empirical risk function $\mathcal{R}_T(\tilde{f}^{(w)}, f)$.

The form of the aggregator will be inferred with the help of the divergence measures introduced above as the solution for $\tilde{f}^{(w)}$ is obtained by selecting a distribution which is as close as possible to the multivariate distribution of f_i's. More specifically, we follow Maasoumi (1986) in generalizing the pairwise criteria of divergence to a general multivariate context:

$$\tilde{D}_\rho(\tilde{f}, f; w) = \sum_{i=1}^{M} w_i \mathcal{R}_{T,\rho}(\tilde{f}, f_i), \tag{1}$$

where

$$\mathcal{R}_{T,\rho}(\tilde{f}, f_i) = \frac{1}{\rho(\rho+1)} \sum_{t=1}^{T} \tilde{f}_t \left[\left(\frac{\tilde{f}_t}{f_{i,t}} \right)^\rho - 1 \right]. \tag{2}$$

$\mathcal{R}_{T,\rho}(\tilde{f}, f_i)$ is the generalized entropy divergence between the aggregator $\tilde{f}$ and each of the prospective models f_i. The aggregator that minimizes $\tilde{D}_\rho(\tilde{f}, f; w)$ is given by

$$\tilde{f}_t^{(w)} \propto \left[\sum_{i=1}^{M} w_i f_{i,t}^{-\rho} \right]^{-1/\rho}. \tag{3}$$

The linear and convex pooling of models are obtained as special cases. The case $\rho = -1/2$ corresponds to our preferred Hellinger distance aggregator.

2.3 Convex Aggregation

The dominant (convex) aggregator of the candidates $\{f_1, ..., f_M\}$ is obtained using the Kullback-Leibler divergence ($\rho = -1$) and is given by

$$f^{(w)} = \sum_{m=1}^{M} w_m f_m,\ w \in \mathcal{W}^M,$$

with its estimator denoted by $\tilde{f}_T^{(w)}$. Model selection is a special case with $w \equiv e_i = (0, 0, ..., 1, 0, ..., 0)$ with $i = 1, ..., M$.

When the density properties of the w are recognized, one may incorporate penalties for departures of the distribution of weights (w) from a priori distributions or desired distributions of weights (π) that may reflect an ordering of the models. For example, consider the linear aggregator $\tilde{f}_w = \sum_{m=1}^{M} w_m f_m$ of an unknown regression function f. Then, the aggregation weights may solve the following penalized optimization problem

$$\min_{w \in \mathcal{W}^M} \left[\sum_{m=1}^{M} w_m \mathcal{R}_T(\tilde{f}_T^{(w)}, f) + \frac{\beta}{T} \mathcal{KL}(w, \pi) \right],$$

where $\beta > 0$ is a penalty parameter, $\mathcal{KL}(w, \pi) = \sum_{m=1}^{M} w_m \ln\left(\frac{w_m}{\pi_m}\right)$ is the Kullback-Leibler divergence between w and π, and $\pi \in \mathcal{W}^M$ is a prior probability density. This could also be a convenient device when M is large relative to T, as in variable selection problems with 'big data' attributes. The solution for the above penalized optimization problem is driven by the form of the entropy divergence function. With the Kullback-Leibler divergence, the aggregation weights take an exponential form

$$w_i^* = \frac{\exp(-T\mathcal{R}_T(\tilde{f}_T^{(w)}, f)/\beta)\pi_i}{\sum_{m=1}^{M} \exp(-T\mathcal{R}_T(\tilde{f}_T^{(w)}, f)/\beta)\pi_m}.$$

Note that this is the quasi-Bayesian approach of Chernozhukov and Hong (2003) where the estimates of w can be obtained using MCMC methods.

2.4 An Illustrative Example: Portfolio Construction

Suppose now that R_i^e denotes the excess return on the risky asset i ($i = 1, ..., N$), P signifies the data generating measure and Q is the risk-neutral measure. An interesting problem to

consider is to find the risk-neutral measure Q with minimal entropy relative to the physical measure P (Stutzer, 1995). The solution to this problem is obtained as

$$Q^* = \underset{Q}{\operatorname{argmin}}\ \mathbb{E}^Q\left[\ln\left(\frac{dQ}{dP}\right)\right]$$

subject to the no-arbitrage restriction

$$\mathbb{E}^Q[R_i^e] \equiv \int R_i^e dQ = 0 \text{ for } i = 1, ..., N.$$

The solution Q^* gives rise to the following density

$$\frac{dQ^*}{dP} = \frac{\exp\left(\sum_{i=1}^N w_i^* R_i^e\right)}{\mathbb{E}\left[\exp\left(\sum_{i=1}^N w_i^* R_i^e\right)\right]},$$

where the density parameters (weights) $w^* = (w_1^*, ..., w_N^*)$ are the solution to the problem

$$w^* = \underset{w=(w_1,...,w_N)}{\operatorname{argmin}}\ \ln\mathbb{E}\left[\exp\left(\sum_{i=1}^N w_i R_i^e\right)\right]$$

One interesting observation is that $\ln\mathbb{E}\left[\exp\left(\sum_{i=1}^N w_i R_i^e\right)\right]$ is the cumulant generating function of $\sum_{i=1}^N w_i R_i^e$ which characterizes all the information in the distribution of the excess returns. When excess returns are assumed to be multivariate normal, all cumulants beyond the first two cumulants are zero and the above optimization problem collapses to the usual mean-variance portfolio problem

$$w^* = \underset{w=(w_1,...,w_N)}{\operatorname{argmin}}\ \mathbb{E}\left[R^e\right]' w + 0.5 w' \mathrm{Cov}[R^e] w$$

with the closed-form solution $w^* = -\mathrm{Cov}[R^e]^{-1}\mathbb{E}\left[R^e\right]$.[2] Substituting for w^*, the relative entropy (Kullback-Leibler) bound becomes $0.5\mathbb{E}\left[R^e\right]'\mathrm{Cov}[R^e]^{-1}\mathbb{E}\left[R^e\right]$ which in the case of

[2]It follows that the vector of relative portfolio weights invested in N risky assets is $\hat{w}_{MV} = -\mathrm{Cov}[R^e]^{-1}\mathbb{E}\left[R^e\right] / \left(1_N \mathrm{Cov}[R^e]^{-1}\mathbb{E}\left[R^e\right]\right)$. A natural benchmark strategy is an equal-weighted portfolio with $\hat{w}_{EW} = 1/N$. Comparing the out-of-sample performance (over the last K observations) of these two strategies via the Sharpe ratio boils down to computing the test statistic $z = (\hat{\sigma}_2\hat{\mu}_1 - \hat{\sigma}_1\hat{\mu}_2)/\hat{\omega}$ with

$$\hat{\omega}^2 = \frac{1}{K}\left(2\hat{\sigma}_1^2\hat{\sigma}_2^2 - 2\hat{\sigma}_1\hat{\sigma}_2\hat{\sigma}_{12}\right) + \frac{1}{2}\hat{\mu}_1^2\hat{\sigma}_2^2 + \frac{1}{2}\hat{\mu}_2^2\hat{\sigma}_1^2 - \frac{\hat{\mu}_1\hat{\mu}_2}{\hat{\sigma}_1\hat{\sigma}_2}\hat{\sigma}_{12}^2,$$

where $\hat{\mu}_1$, $\hat{\mu}_2$, $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}_{12}$ are the estimated sample means, variances and covariances of the two portfolios. Under some assumptions, the statistic z is distributed as a standard normal random variable. DeMiguel, Garlappi and Uppal (2011) do not find sufficient evidence for the dominance of the optimal portfolio weight relative to the naive, equally-weighted, scheme. They attribute this finding to the estimation error that accompanies the construction of the optimal portfolio.

one asset becomes the squared Sharpe ratio of this asset return. It appears that a large part of the entropy is accounted for by the higher than second cumulants which arises from the non-Gaussianity of the excess return data. Thus, ignoring these higher moments in measuring entropy and dependence will result in a significant misspecification and spurious dynamics in the first two moments. Once higher moments and more general loss/risk functions are allowed for, most "anomalies" and "puzzles" tend to diminish in terms of magnitude and economic significance (see Stutzer, 1995, 2016; Ghosh, Julliard and Taylor, 2017; among others).

Note that the assets need not be combined by linear pooling which implicitly assumes a perfect substitutability of the different assets. Suppose that $\tilde{R}$ denotes the aggregator (portfolio). As above, we characterize the solution for $\tilde{R}$ by selecting a distribution which is as close as possible to the multivariate distribution of R_i's using the following measure of divergence:

$$D_\rho(\tilde{R}, R; w) = \sum_{i=1}^{N} w_i \left\{ \sum_{t=1}^{T} \tilde{R}_t \left[\left(\frac{\tilde{R}_t}{R_{i,t}} \right)^\rho - 1 \right] \Big/ \rho(\rho+1) \right\},$$

The aggregator that minimizes $D_\rho(\tilde{R}, R; w)$ subject to $\sum_{i=1}^{N} w_i = 1$ is given by

$$\tilde{R}_t^* \propto \left[\sum_{i=1}^{N} w_i R_{i,t}^{-\rho} \right]^{-1/\rho},$$

with the linear pooling obtained for $\rho = -1$ and the Hellinger distance aggregator obtained for $\rho = -1/2$.

3 Forecast Combination

Suppose now that $f = \{f_1, ..., f_M\}$ denote M forecasts for variable y_{t+1} $(t = 0, ..., T)$ and a convex forecast combination is given by

$$f(w) = \sum_{m=1}^{M} w_m f_m = \mathbf{w}'\mathbf{f},$$

where $\mathbf{w} \in \mathcal{W}^M, \mathcal{W}^M = \left\{ \mathbf{w} \in \mathbb{R}^M : w_m \geq 0, \sum_{m=1}^{M} w_m = 1 \right\}$. The benchmark forecast combination is the equal-weight mixing with $w_m = 1/M$.

The mean-squared forecast errors (MSFE) over K periods are defined as

$$\hat{\sigma}^2(w) = \frac{1}{K} \sum_{t=T-K}^{T} (y_{t+1} - \mathbf{w}'\mathbf{f}_t)^2.$$

One method for forecast combination (Granger-Ramanathan, GR) select the vector of mixing weights $\mathbf{w}$ to minimize MSFE which is equivalent to running the least squares regression

$$y_{t+1} = \mathbf{w}'\mathbf{f}_t + \varepsilon_{t+1}, \ t = T-K, ..., T,$$

with a solution

$$\hat{\mathbf{w}}_{GR} = \left(\sum_{t=T-K}^{T} \mathbf{f}_t \mathbf{f}_t' \right)^{-1} \left(\sum_{t=T-K}^{T} \mathbf{f}_t y_{t+1} \right).$$

But these weights are unconstrained which results in poor forecast performance. The constrained GR forecast weights are obtained (using quadratic programming) as

$$\begin{aligned} \hat{\mathbf{w}}_{CGR} &= \underset{\mathbf{w}}{\text{argmin}} \mathbf{w}'\mathbf{A}\mathbf{w} \\ \text{s.t.} \ \sum\nolimits_{m=1}^{M} w_m &= 1 \text{ and } 0 \leq w_m \leq 1, \end{aligned}$$

where $\mathbf{A} = \sum_t \mathbf{e}_{t+1}\mathbf{e}_{t+1}'$ and $\mathbf{e}_{t+1} = y_{t+1} - \mathbf{f}_t$. A special case for this method (Bates-Granger) is obtained under the assumption that $\mathbf{A}$ is diagonal with weights given by $\hat{w}_{m,BG} = \hat{\sigma}_m^2 / \sum_{j=1}^{M} \hat{\sigma}_j^2$. If the different forecast error variances are approximately equal – which is often the case in practice – this forecast combination is similar to the constant, equal-weight mixing. The weights could also be obtained using a cross-validation criterion and leave-one-out estimator. Finally, other popular combination scheme include the Bayesian model averaging and the Mallows model averaging (Hansen, 2007, 2008).

Instead of relying on the convex aggregation, we turn again to an aggregator of the constant-elasticity-of-substitution (CES) form

$$\tilde{f}_t = \left[\sum\nolimits_{i=1}^{M} w_i f_{i,t}^{-\rho} \right]^{-1/\rho}. \tag{4}$$

Again, our preferred aggregator is based on the value $\rho = -1/2$ which corresponds to the Hellinger distance measure

$$\mathcal{H}(P, Q) = \frac{1}{2} \int \left(p^{1/2} - q^{1/2} \right)^2 dv,$$

where p be the density of some favored benchmark ("pivot"), and q the density of the aggregator $\tilde{f}_t^* = \left[\sum_{i=1}^M w_i f_{i,t}^{1/2}\right]^2$. The mixing weights are obtained by minimizing the above distance measure[3] with respect to w, subject to $w_i \geq 0$ and $\sum_{i=1}^M w_i = 1$. Unlike the other measures in the Cressie-Read divergence family, the Hellinger distance is a proper measure of distance since it is positive, symmetric and it satisfies the triangle inequality. We follow this estimation strategy in the empirical section for forecasting inflation below.

Finally, we would like to remark briefly on some further extensions. First, we should note that all of these forecast combination methods are designed to produce point forecasts. One general method for incorporating the forecast uncertainty and constructing forecast intervals in a model-free way is the conformal predictive inference (Vovk, Gammerman, and Shafer, 2005; Lei, G'Sell, Rinaldo, Tibshirani, and Wasserman, 2018; Chernozhukov, Wüthrich and Yinchu, 2018). Second, this entropy-based approach can be readily adapted to aggregation of density forecasts from models or from experts. For example, one could be interested in combining and summarizing the information from M probability density (mass) forecasts for variable y by M survey participants over N prespecified bins of possible values for y. For details on this, see Gospodinov and Maasoumi (2019).

3.1 Bregman Pseudo-Distances for Forecast Evaluation

Forecast evaluation requires a choice of a loss function. One flexible class of loss functions are the Bregman (1967) pseudo-distances

$$B_\phi(p, q) = \phi(p) - \phi(q) - \phi'(q)(p - q),$$

where $\phi : \mathbb{R} \to [0, +\infty)$ is again a convex, continuously differentiable function, $p = dP/d\mu$ and $q = dQ/d\mu$. One interesting result (Stummer and Vajda, 2011) is that the scaled Bregman pseudo-distances are equal to the divergence measures considered above

$$B_\phi(P, Q|Q) = D_\phi(P, Q).$$

Two popular choices of Bregman pseudo-distances (Patton, 2020) for forecast evaluation (forecast f for variable y) are based on $\phi(x; k) = |x|^k$, $k > 1$ (homogeneous Bregman loss):

$$L(y, f; k) = |y|^k - |f|^k - k \operatorname{sgn}(f)|f|^{k-1}(y - f)$$

[3]Densities p and q are estimated by a kernel density estimator and the integral is evaluated numerically.

or based on $\phi(x;a) = 2a^{-2}\exp(ax)$, $a \neq 0$ (non-homogeneous Bregman loss):

$$L(y,f;a) = \frac{2}{a^2}\left[\exp(ay) - \exp(af)\right] - \frac{2}{a}\exp(af)(y-f).$$

Different choices of a and k give rise to different penalties for over- and under-predictions. The standard mean squared error (MSE) is obtained for $k = 2$ and $a \to 0$.

3.2 Forecasting Inflation

We will illustrate the advantages of the proposed aggregation approach for forecasting U.S. core inflation (CPI less food and energy) for the period 1988:01–2018:04. The underlying data is monthly year-over-year inflation rate and the forecasts are 12-month ahead forecasts. We consider 5 individual models for inflation. One model is concerned with domestic slack (PC: Phillips curve), another focuses on the forward-looking component of commodity prices (CY: convenience yield model (Gospodinov and Ng, 2013; Gospodinov, 2016), and a third model is completely statistical (MA: integrated moving average model (1,1) model (Stock and Watson, 2007)). As a benchmark model, we use a simple historical average (HA) model. Survey expectations constitute another useful source of information about future inflation. Since these are model-free forecasts, we use it (BC: Blue Chip survey of expected CPI inflation) as a pivot in constructing

Given the fundamental uncertainty surrounding the underlying data generating process for inflation, it is not unreasonable to assume that all models for describing and forecasting the inflation dynamics are inherently incomplete as they are designed to capture different features of interest. Since the "true" model is unlikely to be in any set of candidate models, reliance on a single model for policy analysis or forecasting is sub-optimal and results in loss of information. Using the aggregation approach (AG), we average forecasts from several candidate models (PC, CY, MA and HA), where the mixing weights are estimated by shrinking the aggregator towards survey expectations. The average has a constant-elasticity-of-substitution form that relaxes the assumption that the candidate models are perfectly substitutable - which is implicit in the linear pooling of forecasts.

The individual model parameters are estimated using recursive model estimation (initial sample: 1988:01–1996:12), while the aggregation weights are estimated over a separate training sample (initial sample: 1997:01–2001:12). The pseudo out-of-sample evaluation is

over the period 2002:01-2018:04. Results for the forecast performance of the different models, based on the two Bregman loss functions, are presented in Table 1. The value for the aggregator (AG) is standardized to be equal to one. As a result, numbers larger than one indicate that the corresponding model is dominated by AG.

Table 1. Bregman loss functions for different forecasting models

	PC	HA	MA	BC	CY	AG
Homogeneous Bregman Loss ($k > 1$)						
$k = 1.1$	2.2785	1.9222	2.2682	1.6576	1.3662	1.0000
$k = 2$ (MSE)	1.9290	2.0506	2.2673	1.7682	1.6063	1.0000
$k = 3$	1.8157	2.1712	2.2651	1.9128	1.9489	1.0000
$k = 3.5$	1.8088	2.2210	2.2642	1.9931	2.1540	1.0000
$k = 4$	1.8212	2.2633	2.2638	2.0783	2.3843	1.0000
Non-homogeneous (exponential) Bregman Loss ($a \neq 0$)						
$a = -1$	2.3653	1.8120	2.2710	1.5592	1.1570	1.0000
$a = -0.5$	2.0824	1.9354	2.2695	1.6514	1.3440	1.0000
$a \to 0$ (MSE)	1.9289	2.0506	2.2673	1.7683	1.6063	1.0000
$a = 0.5$	1.8689	2.1497	2.2658	1.9122	1.9702	1.0000
$a = 1$	1.8773	2.2258	2.2658	2.0858	2.4728	1.0000

Notes: All losses are expressed as ratios to that of the aggregator (AG) model.

The results in Table 1 show that the entropy-based forecast combination dominates individual model forecasts across all loss functions. The aggregation approach reduces the mean square forecast error by more than 60% for individual models (including survey forecasts) with the forecast gains being even larger for asymmetric loss functions that penalize more heavily over-predictions than under-predictions. Given the challenges in forecasting inflation, these are huge improvements. For the individual models, BC and CY work best except when over-predictions are very costly. The forecast average assigns the largest weight to the commodity model but balances the forward-looking, yet volatile, nature of these forecasts with the more stable behavior of the historical average and survey forecasts. The time variation of the mixing weights also reveals interesting information about the relative importance of the individual models over this historical period. Unlike the individual forecasts, the aggregation approach produces an unbiased forecast with a Mincer-Zarnowitz regression slope coefficient of 1.062 (0.234). "Intercept corrections" (à la Klein/Theil) can lead to further forecast improvements.

4 Policy Implications

Decision makers routinely employ multiple models to interpret empirical evidence and formulate responses to shocks and policy scenarios, acknowledging that these models are only partial, low-resolution maps of the underlying economic environment. This paper highlights the role of model ambiguity and uncertainty and advances the idea of model and forecast aggregation as a robust analytical framework for accounting for the incomplete nature of these models. In this respect, it shares some commonalities – in terms of confronting this uncertainty – with the robust control approach, articulated by Hansen and Sargent (2001) and in a series of related papers (see also Karantounias, 2020). It is important to stress that in our framework, we dispense the notion of a true model and treat the candidate models as genuinely misspecified for the latent object of interest. This stands in contrast to some of the prevailing approaches, with important implications for designing robust policies.

The most common perspective, that includes Bayesian model averaging and model selection, is conditioned on one of the models in the decision-maker dictionary being 'true'. In this approach, the ambiguity about the true model is resolved asymptotically, and the mixture that summarizes the beliefs about the individual models assigns a unity weight to one of the models. Another possibility is to partially relax this assumption and allow the unknown true model to belong to a neighborhood of an approximate reference model. While this 'model ambiguity' approach accounts for some of the uncertainty around the reference model, the policy formulation still relies on information embedded in a single model. A third possibility is to assume that a true model exists but it is too complicated or cumbersome to implement (see Bernardo and Smith, 1994). For all practical purposes, this coincides with our setting as all of the candidate models should be viewed as approximations of this fully-specified belief model and hence inherently misspecified. Some important insights about the differences in policy implications for the latter approach can be gleaned from the analysis in Acemoglu, Chernozhukov and Yildiz (2006). If the uncertainty across the models, entertained by the decision makers to interpret an information signal, is not resolved asymptotically but it persists, the decision makers may exhibit a persistent divergence of opinions, even after observing the same infinite sequence of signals. This example illustrates the complex nature of decision making under uncertainty and the potential benefits of optimal aggregation.

References

Acemoglu, D., V. Chernozhukov, and M. Yildiz (2006): "Learning and disagreement in an uncertain world," MIT Department of Economics Working Paper No. 06-28.

Bernardo, J., and A. Smith (1994): *Bayesian Theory*, Wiley.

Bregman, L. M. (1967): "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics* 7, 200–217.

Csiszár, I. (1972): "A class of measures of informativity of observation channels," *Periodica Mathematica Hungarica* 2, 191–213.

Chernozhukov, V., and H. Hong (2003): "An MCMC approach to classical estimation," *Journal of Econometrics* 115, 293–346.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2018): "Exact and robust conformal inference methods for predictive machine learning with dependent data," Proceedings of the 31st Conference On Learning Theory, *PMLR* 75, 732–749.

Cressie, N., and T. Read (1984): "Multinomial goodness of fit tests," *Journal of the Royal Statistical Society B* 46, 440–464.

Ghosh, A., C. Julliard, and A. P. Taylor (2017): "What is the consumption-CAPM missing? An information-theoretic framework for the analysis of asset pricing models," *Review of Financial Studies* 30, 442–504.

Gospodinov, N. (2016): "The role of commodity prices in forecasting U.S. core inflation," Atlanta Fed Working Paper 2016-5.

Gospodinov, N. and E. Maasoumi (2021): "Generalized aggregation of misspecified models: With an application to asset pricing," *Journal of Econometrics* 222, 451–467.

Gospodinov, N. and E. Maasoumi (2019): "Aggregating probability distributions from experts," Manuscript.

Gospodinov, N. and E. Maasoumi (2025): "Ethical price indices," Manuscript.

Gospodinov, N., and S. Ng (2013): "Commodity prices, convenience yields, and inflation," *Review of Economics and Statistics* 95, 206–219.

Hansen, B. E. (2007): "Least squares model averaging," *Econometrica* 75, 1175–1189.

Hansen, B. E. (2008): "Least squares forecast averaging," *Journal of Econometrics* 146, 342–350.

Hansen, L. P., and T. J. Sargent (2001): "Robust control and model uncertainty," *American Economic Review* 91, 60–66.

Karantounias, A. G. (2020): "Model uncertainty and policy design," Atlanta Fed Policy Hub Paper 2020–17.

Kobus, M., M. Kapera, and E. Maasoumi (2024): "Generalized multivariate gaps: A proposed gender gap," The University of Chicago Stone Center Working Paper Series Paper No. 25-03.

Lei, J., M. G'Sell, A. Rinaldo, R. J. Tibshirani, and L. Wasserman (2018): "Distribution-free predictive inference for regression," *Journal of the American Statistical Association* 113, 1094–1111.

Maasoumi, E. (1986): "The measurement and decomposition of multi-dimensional inequality," *Econometrica* 54, 991–997.

Maasoumi, E. (1993): "A compendium to information theory in economics and econometrics," *Econometric Reviews* 12, 1–49.

Mudekereza, F. (2025): "Robust social planning," Working paper, MIT.

Patton, A. (2020): "Comparing possibly misspecified forecasts," Journal of Business & Economic Statistics 38, 796–809.

Stock, J. H., and M. W. Watson (2007): "Why has inflation become harder to forecast?," *Journal of Money, Credit and Banking* 39, 3–33.

Stummer, W., and I. Vajda (2012): "On Bregman distances and divergences of probability measures," *IEEE Transactions on Information Theory* 58, 1277–1288.

Stutzer, M. (1995): "A Bayesian approach to diagnosis of asset pricing models," *Journal of Econometrics* 68, 367–397.

Stutzer, M. (2016): "Entropic diagnostics for asset pricing SDFs: A critique," Working paper.

Vovk, V., A. Gammerman, and G. Shafer (2005): *Algorithmic Learning in a Random World*, New York: Springer.