MEASURING FORECASTER ATTENTION

By Susannah Scanlan

I propose a novel method to measure forecaster attention to different dimensions of information. I do this by establishing a connection between rational inattention models and econometric factor models. I show that a theoretical rationally inattentive forecaster's choice of what to pay attention to will be reflected in the lower dimensional structure of their implied multivariate forecasts. With this mapping, I propose an estimator to measure forecaster attention to different dimensions of information. I find that professional forecasters pay less attention to less important sources of information and worse performing forecasters display less attention overall, despite sharing similar reduced-form forecasting models. I find that the relationship between attention and the cost-to-benefit ratio of a dimension of information is, in general, nonlinear, in contrast to what is implied by the standard assumption of Mutual Information. I illustrate the nonlinearity in attention across professional forecasters by measuring attention to information associated with long-term inflation under low-, medium- and high-inflation regimes.

0. INTRODUCTION

How forecasters make predictions, and in particular how they decide what information to pay attention to, is an important question for policymakers. A measure of forecaster attention would be valuable not only as an indicator of risk, as attention can shift rapidly in different economic regimes, but also as a tool for understanding the rules which govern how individuals allocate their attention across information sources. In this paper, I propose a framework to implicitly measure forecaster attention to different dimensions of information from panel forecast datasets.

In the context of making predictions for multiple variables, how a forecaster decides to allocate their attention will determine the strength of correlations across the forecasted

---

Email: ss5605@columbia.edu. I would like to thank Michael Woodford, Serena Ng, Hassan Afrouzi, Mark Dean, Bernard Salanié, Simon Lee, Jushan Bai, and Jennifer L'aO for their helpful guidance and comments. I additionally would like to thank the members of Columbia's Decision and Cognition Lab. This work has benefited from the feedback of seminar participants at the Federal Reserve Bank of New York, the Federal Reserve Board of Governors, and participants of the SED 2024 Conference. This paper was previously posted as "Attention Allocation and the Factor Structure of Forecasts". An earlier version of the paper appears in Susannah Scanlan's PhD dissertation. The views in this paper are my own and do not represent those of any institution.

variables. Consider a forecaster making predictions of interest rates, residential investment, and consumer spending. Very likely, a large part of the forecasts for residential investment and consumer spending will be captured by the predictions for interest rates. If a forecaster pays more attention to and acquires more information on interest rates, their predictions for residential investment and consumer spending will be more correlated along the "interest rate" dimension of information (factor). Additionally, because of the relevance of interest rate information to all three of the forecasted variables, it would be optimal for the forecaster to pay more attention to factors that capture variations in interest rates rather than to housing- or consumption-specific factors.

I propose a measure of forecaster attention based on the relative strength of the underlying correlation, or factor, structure. The measure of attention is grounded in a theoretical model in which I map a forecaster's choice of how much attention to allocate to different dimensions of information to the factor decomposition of implied forecasts. I additionally propose a new and generalized class of cost functions — parametric models which govern how individuals optimally decide to acquire information — for which the mapping from attention to factor structure, and thus measure of forecaster attention, is valid.

With the empirical measures of professional forecaster attention, I estimate what type of cost function best represents the decision of forecasters. The cost function captures how attention changes given the importance of a dimension of information, and can take different functional forms. I find that the most popular assumption of an information cost proportional to mutual information, as in Sims (2003), is only a reasonably assumption for top-performing forecasters. As forecast performance decline, the estimated form of cost function moves toward the Fisher information cost function of Hébert and Woodford (2021).[1] The form of an economic agent's cost function can have important general equilibrium consequences. Hébert and La'O (2023), for example, show that Fisher information costs can lead to nonfundamental volatility in generalized beauty contest games, whereas mutual information does not.

A forecaster's degree of attentiveness will affect the strength of the underlying latent factor structure, captured mathematically by the size of the eigenvalues from a singular value

[1] Fisher information is the limiting case of Hébert and Woodford's proposed neighborhood-based cost function in the case of a linear quadratic Gaussian tracking problem.

decomposition of the time series forecast panel. Relative to someone who can fully process all relevant information, the eigenvalues of a forecast made by someone who faces a cost on acquiring and processing information will be shrunken in size.[2] The degree of shrinkage on an eigenvalue captures the amount of attention a forecaster pays to the dimension of information aligned with the corresponding eigenvector.

In order to measure relative forecaster attention, I propose an estimator for the relative shrinkage along the eigenvalues of forecasts. This estimator is based on a two-step procedure and requires a proxy full information forecast. Attention is measured along the eigenvectors from a singular value decomposition of the full information forecast proxy, and the shrinkage estimator measures the size of rationally inattentive eigenvalues in relation to the full information eigenvalues. As a full information proxy I use both the Consensus or median forecast from the *Survey of Professional Forecasters*, as well as the Federal Reserve Greenbook Forecast.

The mapping from attention to the factor structure of forecasts, and hench the measure, is valid for a new and generalized set of cost functions, which I call separable spectral. These cost functions are defined as additively separable functions over the eigenvalues of the prior and posterior variance matrix — prior being the forecasters belief about the state of the world before acquiring information and posterior being their belief after acquiring information. Intuitively, separable spectral information costs add up the cost of reducing uncertainty along each dimension of information spanned by the variance matrices. Separable spectral cost functions are defined only for rational inattention models that fall under the multivariate linear quadratic Gaussian tracking problem. The connection between the theoretical attention allocation for a forecaster who faces a cost on processing information and the eigenvalue shrinkage in the corresponding forecasts is valid for separable spectral information cost functions.

Separable spectral costs nest existing popular cost functions, including Mutual Information, or a reduction in Shannon entropy, and Fisher information. Mutual information is a tractable way to model information acquisition, likely contributing to its popularity.[3]

[2] See, for example, Matějka and McKay (2015).

[3] Techniques involving mutual information were developed in the signal processing literature (Cover and Thomas (2006)). Examples of papers that use a mutual information cost function include: Sims (1998), Sims

However, mutual information also has restrictive properties on how individuals process information that are contradicted by observed experimental evidence.[4] Mutual information does not permit differences in acquisition costs stemming from differences in perceptual distance, such as the difference between two numbers, but the magnitude of these differences have been shown to be important in how people process information (Khaw et al. (2020)). A new neighborhood-based cost function, which addresses this concern, was proposed by Hébert and Woodford (2021). It is based on the notion that some states are easier to distinguish than others. In the case of a linear quadratic Gaussian tracking problem, the neighborhood-based cost function takes the form of a cost on Fisher information. As noted previously, differences in the type of cost functions faced by agents can lead to different general equilibrium results.

The exact form of information cost function determines how attention changes with the relative importance of a source of information. In particular, mutual information implies a linear relationship between attention and the cost-to-benefit ratio of a source of information. Fisher information indicates nonlinear relationship between attention and the cost-to-benefit ratio, with a declining marginal change in attention as information becomes less valuable. Under Fisher information, forecasters pay relatively more attention to more important dimensions of information, whereas under Mutual Information there is a linear relationship between attention and the information cost-to-benefit ratio.

I estimate the form of information cost function that best describes forecaster behavior. I use data from the *Survey of Professional Forecasters*, and I measure forecaster attention to the Consensus, or median, forecast. I estimate the parametric form of the cost function for top-, medium-, and poorly-performing forecasters by minimum distance. I find that the estimated linear cost of information is higher for worse-performing forecasters and that the estimated functional form of the cost function changes with forecast performance. Top-performing forecasters have estimated cost functions closer to that of mutual information,

---

(2003), Mackowiak and Wiederholt (2009), Woodford (2009), Mondria (2010), Matějka and McKay (2015), Matějka (2015) Yang (2015), Bartoš et al. (2016), Steiner et al. (2017), Fulton (2017), Caplin et al. (2019), Kőszegi and Matějka (2020), Afrouzi and Yang (2021).

[4] In particular, mutual information indicates that an agent's attention should only be determined by the relevant payoff, but Dean and Neligh (Forthcoming) present experimental evidence that shows subjects behave differently if states are more difficult to distinguish, despite the payoffs remaining fixed.

whereas bottom-performing forecasters have cost functions close to Fisher information. Medium-performing forecasters fall in between.

In addition to analyzing the form of forecaster cost function, the attention measure can also shed light on the form of forecaster model, as well as forecaster attention to economically-meaningful factors. With the estimator, I can examine how rotated the eigenvectors from a rationally inattentive forecast are from the eigenvectors of a full information forecast. I find minimal rotation, indicating that professional forecasters have the same reduced-form forecast model, and instead differ on the intensity of their relative attention.

In order to illustrate the differences between linear and nonlinear shifts in attention, I consider forecaster attention to long-term inflation under different inflationary regimes — a subject of particular interest to policymakers.[5] The attention of middle- and bottom-performing forecasters displays little difference in between the low- and medium-inflation regimes (2010Q1-2019Q4 and 1985Q1-2007Q4, respectively), but increases significantly following the 2021 inflation surge. The attention of top-performing forecasters displays notable shifts in between each of the regimes. This finding is consistent with the theoretical model that implies forecasters pay more attention to sources of information for which they are more uncertain.[6]

This paper is related to the literature on rational inattention models and factor models. Within rational inattention models, the paper is most related to research on rational inattention models in the linear quadratic Gaussian setting (Sims (2003), Kőszegi and Matějka (2020); Afrouzi and Yang (2021); Fulton (2017)). Unlike many papers on expectations formation (Mankiw et al. (2003); Coibion and Gorodnichenko (2015)), I focus on a multivariate forecast with a multidimensional state space. In the linear quadratic Gaussian context, this has been modeled (Afrouzi and Yang (2021)), but the factor structure implications in a panel dataset have not been fully explored. Relatedly, Dovern (2015), Andrade et al. (2016), and Herbst and Winkler (2021) consider the factor structure of disagreement.

Rationally inattentive decision-making induces a lower-dimensional factor structure in implied data. The decision of a rationally inattentive forecaster is similar to that of a Now-

[5] Measuring attention to economically-meaningful factors, such as long-term inflation, requires an additional assumption of how forecasts are formed.

[6] Afrouzi and Yang (2021) examine attention under different prior uncertainties.

caster, as in the dynamic factor model setup of Giannone et al. (2008).[7] The key difference is that the observation error, and hence Kalman gain on higher-frequency signals, arises from the information choice of the forecaster. A rationally inattentive forecaster has the choice of acquiring more signals to get a more precise estimate of the latent economic states. The endogenous information choice leads to predictions about the intensity of attention along more and less important dimensions. Previous papers have also pointed out shrinkage and thresholding in rational inattention models (Afrouzi and Yang (2021); Kőszegi and Matějka (2020)), albeit under terms not typically associated with the factor structure of panel data. By drawing this connection and formalizing it, I am able to analyze the multivariate attention of forecasters using tools from factor analysis.

In Section 1, I define separable spectral information costs and present the forecaster's optimal choice of information acquisition. In Section 2, I analyze optimal attention through the lens of factor analysis and present an estimator for forecaster attention. In Section 3, I present empirical results.

1. A MODEL OF RATIONAL INATTENTION WITH SEPARABLE SPECTRAL INFORMATION COSTS

In order to capture the optimal behavior for a broad set of possible information acquisition processes, I define Separable Spectral information costs - a novel and generalized class of information costs in the context of the linear quadratic Gaussian tracking problem. The mapping from optimal attention allocation to shrinkage in the factor structure of implied forecasts is valid for all Separable Spectral cost functions, thus permitting a later estimation of the exact form of the forecaster cost function. The optimal attention of a forecaster that comes out of the optimization problem presented in this section serves as the data generating process of the factor model presented subsequently.

1.1. *Model Preliminaries and Objective*

The forecaster's goal is to minimize their forecast error. They make predictions for n variables. The true realization of the variables is denoted by X and the agent's forecast is denoted $\hat{X}$. Before making a forecast, agents get to choose an information set $\mathcal{I}$ which is subject to a cost $C(\mathcal{I})$. More precise information helps forecasters make more accurate

[7] This connection was noted in Herbst and Winkler (2021).

forecasts. The agent's objective function is as follows:

$$\min_{\mathcal{I}} \mathbb{E}\left[\min_{\hat{X}} \frac{1}{2}\mathbb{E}\left[(\hat{X}-X)'W(\hat{X}-X)+\omega C\left(\mathcal{I}\right)|\mathcal{I}\right]\right]$$

The variables of interest linearly depend on a set of latent states, $\tilde{\mathbf{a}}$ with variance P. The forecaster has prior knowledge of P, so the shocks can be normalized to have unit variance.

$$\underset{(n\times 1)}{X} = \underset{(n\times r)}{\tilde{B}'}\underset{(r\times 1)}{\tilde{\mathbf{a}}} + \underset{(n\times 1)}{u} \qquad r \leq n$$

$$\tilde{\mathbf{a}} = P^{\frac{1}{2}}\mathbf{a} \qquad \mathbf{a} \sim N\left(0, \mathbb{I}\right)$$

$$u_i \sim N\left(0, \sigma_{ui}^2\right), \sigma_{ui}^2 << \text{eig}\left(P\right)_i \qquad \text{for } i = 1, \ldots, n$$

The latent states ($\mathbf{a} = (a_1 \ldots a_r)'$) is potentially observable, whereas the $n \times 1$ vector u is unobservable. The variance of the unforecastable component, σ_{ui}^2, is much smaller than that of the forecastable component, characterized by i^{th} eigenvalue of the matrix P. Both processes are Gaussian. Importantly, the dimension of the state vector, r, can be equal to the number of forecasted variables, n. Without a cost function it could be optimal for agents to acquire idiosyncratic signals for each of the forecasted variables.

If there were no cost on acquiring information (e.g. $C(\mathcal{I}) = 0 \; \forall \mathcal{I}$), the agent would acquire perfect signals of $\mathbf{a}$. I call this forecast $\hat{X}^*$ and it represents the full information rational expectations forecast:

$$\hat{X}^* = \tilde{B}' P^{\frac{1}{2}}\mathbf{a}$$

Informationally-constrained forecasters differ from full information forecasters in that their forecasts are based on their best guess ($\hat{\mathbf{a}}$) of the potentially observable component of the state instead of the true realization.

$$\hat{X} = \tilde{B}' P^{\frac{1}{2}}\hat{\mathbf{a}}$$

The diagonal weight matrix W, of dimension $n \times n$, in the forecaster's objective function captures how much a forecaster cares about the forecast error of a specific variable. A forecaster might care a lot more about ensuring the accuracy of one of their forecasts (next quarter's GDP growth) over another (two-year ahead residential investment). Higher diagonal value of W indicate the forecaster cares more about their forecast for that variable.

Forecasters are restricted from having a negative weight (value getting a forecast wrong), thus the weight matrix can be split and applied to the forecasted value and actual realization:

$$(\hat{X} - X)'W(\hat{X} - X) \Rightarrow (W^{\frac{1}{2}}\hat{X} - W^{\frac{1}{2}}X)'(W^{\frac{1}{2}}\hat{X} - W^{\frac{1}{2}}X)$$

Both the weight matrix, W and the general variance matrix, P, can be subsumed into the coefficient matrix $\tilde{B}$. Define a new coefficient matrix, B, that captures the weight a forecaster puts on the accuracy of their predictions for different variables, variance of underlying economic states, along with the more standard correlation between the latent states and the variables being forecasted.

$$B = W^{\frac{1}{2}}\tilde{B}P^{\frac{1}{2}}$$

The forecaster's best guess of the latent states are based on their choice of signal (s) — specifically the precision of the signal (captured by Q). Forecasters are restricted only to choosing signals which are linear combinations of the latent states with Gaussian error:

$$\underbrace{s}_{m\times 1} = \underbrace{Z}_{m\times r}\underbrace{\mathbf{a}}_{r\times 1} + \underbrace{\epsilon}_{m\times 1} \qquad\qquad \epsilon \sim N(0, Q)$$

Forecasters get to choose both the signal structure Z and signal error variance Q. After seeing the signal realization, agents update their beliefs about the realization of the state $\hat{\mathbf{a}}$ according to Bayes' formula.

$$\hat{\mathbf{a}} = Z\left(ZZ' + Q\right)^{-1}s$$

Their posterior variance belief $\hat{P} = \text{Var}(\mathbf{a}|s)$, $\hat{P} = \mathbb{E}\left[(\hat{\mathbf{a}} - \mathbf{a})'(\hat{\mathbf{a}} - \mathbf{a})\right]$, is given by the following:

$$\hat{P} = \mathbb{I} - Z'\left(ZZ' + Q\right)^{-1}Z$$

The point of this algebraic exercise is to emphasize that the quadratic loss part of the objective function can be simplified into exclusively a choice of the expected posterior variance ($\hat{P}$).

$$\mathbb{E}\left[(\hat{X} - X)'W(\hat{X} - X)\right] = \text{tr}\left(BB'\hat{P}\right) + \sum_{i=1}^{n}\sigma_{ui}^2$$

The coefficient matrix B captures three elements of the forecast: the weight forecasters put on different variables, the variance of the underlying shocks, and the standard correlation between shocks and forecasted variables. The sum of the variances of the unforcastable component of the shocks ($\sum_{i=1}^{n} \sigma_{ui}^2$) does not affect the optimal attention allocation and so is dropped from subsequent derivations.

1.2. *Model Costs: Separable Spectral Information Costs*

The forecast error minimization part of the forecaster's optimization problem can be written in terms of a choice of posterior variance. In this section I propose a new class of cost functions, separable spectral information costs, that can also be written in terms of a choice of posterior variance. Thus the choice of what values to choose for a vector of forecasts boils down to what posterior variance the forecaster wants to target. Separable spectral information costs subsume the well-known Mutual Information and more recently proposed Fisher information cost functions. Under this new type of cost function, information costs are additively separable functions of the eigenvalues (hence, "Spectral") of the forecaster's prior and posterior variance matrices.

DEFINITION 1: ***Separable Spectral Cost Function***. *For a positive definite symmetric matrix with spectral decomposition* $Y = V\Lambda V'$ *and* $\Lambda = dg\left(\lambda_1 \dots \lambda_r\right)$*, the monotone and convex function* $\mathbf{f} : \mathbb{R}^{r\times r} \to \mathbb{R}$ *satisfies the following:*

$$\mathbf{f}\left(Y\right) = \sum_{i=1}^{r} f\left(\lambda_i\right) \tag{1}$$

Additionally:

$$As\ \lambda_i \to 0\ f\left(\lambda_i\right) \to \infty$$

The separable spectral cost function is defined as the difference between the mapping of the prior (P) *and posterior* ($\hat{P}$) *variance beliefs by function* $\mathbf{f}$*:*

$$C^{SS}\left(\hat{P}, P\right) = \mathbf{f}\left(\hat{P}\right) - \mathbf{f}\left(P\right)$$

Separable spectral costs subsume existing cost functions, including mutual information and Fisher information.

ASSUMPTION 1: *Forecasters have a cost of information that is characterized by a separable spectral cost function.*

The derivations that follow are valid for forecasters with separable spectral information costs. A longer description of separable spectral information costs can be found in the Appendix.

1.3. *Model Solution: Optimal Attention Allocation and Forecast*

A forecaster will pay more attention to more important information, and in optimum will ignore some sources of less valuable information. The amount of attention they allocate to a dimension of information will reflect the ratio of their own cost of acquiring information to the benefit of the dimension of information.

PROPOSITION 1: *Let $\lambda_1^B, \dots \lambda_r^B$ denote the eigenvalues of BB' ordered from largest to smallest. A the singular value decomposition of the posterior variance matrix yields $\hat{P} = UDU'$, where d_i is the i^{th} eigenvalue and diagonal element of D, and u_i the i^{th} eigenvector and column of U. The informationally-constrained forecasters objective function can be written in terms of a choice of eigenvalues and eigenvectors (d_i and u_i for $i \in 1, \dots r$) of the posterior variance matrix:*

$$\min_{\{d_i, u_i\}_i^r} \sum_{i=1}^{r} d_i u_i' BB' u_i + \omega \sum_{i=1}^{r} f(d_i)$$
$$\text{s.t. } 0 \leq d_i \leq 1$$

An agent will choose an optimal posterior variance with eigenvalues (Ω_i, $i \in 1 \dots r$) that satisfy $\frac{\lambda_i^B}{\omega} = -f'(\Omega_i)$, with the constraint that $0 \leq \Omega_i \leq 1$.

$$\Omega_i = \begin{cases} f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right) & 0 \leq f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right) \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{2}$$

The optimal eigenvectors of the posterior variance are equivalent to the eigenvectors of the matrix BB', denoted U^B.

The proof of Proposition 1 can be found in the Appendix.

The forecaster's optimal choice of posterior variance is: $\hat{P}^* = U^B \Omega U^{B\prime}$. The matrix U^B characterizes the eigenvectors of BB' in descending order of eigenvalue size and Ω is a diagonal matrix with diagonal elements $\Omega_1, \ldots \Omega_r$ corresponding to Proposition 1.

A rationally inattentive forecast will choose a forecast according to their posterior belief following the standard Kalman updating equations, which in turn reflects their choice of posterior variance. More details can be found in the Appendix. Let $\xi = Z' \left(ZZ' + Q\right)^{-1} \epsilon$. The optimal choice of forecast will be:

$$\hat{X} = B'\hat{\mathbf{a}}$$

$$\hat{X} = B' \left(\mathbb{I} - U^B \Omega U^{B\prime}\right) \mathbf{a} + B' \left(\mathbb{I} - U^B \Omega U^{B\prime}\right) \xi$$

The forecaster's optimal choice of attention along each dimension of information (e.g. real activity, inflation) reflects a cost-to-benefit ratio. The cost is fixed for a forecaster, whereas the benefit reflects the importance of the dimension of information. The matrix B, $B = W^{\frac{1}{2}} \tilde{B} P^{\frac{1}{2}}$, consists of three matrices that capture importance: the weight that forecasters put on accuracy for each variable (W), the correlation between latent states and forecasted variables ($\tilde{B}$), and the prior variance of latent states (P). The dimensions along which forecasters allocate attention are captured by the eigenvectors of this aggregate B matrix. The relative importance of a dimension is captured by the eigenvalues of the B matrix. The value λ_i^B is increasing in the value of the prior variance along dimension i, and similarily increasing along dimensions of correlated forecasted variables with high values of the weight matrix W. If forecasters have a higher prior variance belief, or are more uncertain, then the benefit of acquiring information along that dimension will be more valuable. In either case, a higher value of λ_i^B opens the interior solution space in which forecasters allocate more attention.

1.4. *Optimal Attention Under Examples of Separable Spectral Information Costs*

The mapping from a forecaster's optimal choice of attention allocation to the factor structure of their implied forecasts is valid for any information cost structure that can be characterized as Separable Spectral. This includes both mutual information, as well as Fisher information. As described in Proposition 1, a rationally inattentive forecaster with linear information cost ω and a separable spectral information cost function will choose a posterior

variance with eigenvalues that satisfy:

$$\frac{\lambda_i^B}{\omega} = -f'(\Omega_i)$$

In the case of mutual information, the functional forms of $f(\cdot)$ and $f'(\cdot)$, denoted with a subscript "MI", are[8] :

$$f_{MI}(\Omega_i) = -\log(d_i) \qquad f'_{MI}(\Omega_i) = -\frac{1}{\Omega_i}$$

Thus a forecaster's optimal choice of posterior variance eigenvalue along the i^{th} dimension is the cost to benefit ratio of acquiring information:

$$\Omega_i^{MI} = \frac{\omega}{\lambda_i^B}$$

A more general form of separable spectral cost function, denoted with subscript "p", takes the form:

$$f_p(\Omega_i) = \frac{1}{p}(\Omega_i)^{-p} \qquad f'_p(\Omega_i) = -\Omega_i^{-(1+p)}$$

$$\text{for } p \geq 0$$

Under this form, forecasters choose the following posterior variance eigenvalues:

$$\Omega_i^p = \left(\frac{\omega}{\lambda_i^B}\right)^{\frac{1}{1+p}}$$

The case of $p = 1$ corresponds with Fisher information, or neighborhood-based cost functions in a linear quadratic Gaussian setting. For both mutual information and Fisher information the choice of posterior variance eigenvalues can be written in the parametric form:

$$\Omega_i = \left(\frac{\omega}{\lambda_i^B}\right)^{\alpha}$$

[8] As in Afrouzi and Yang (2021), a linear cost on mutual information is characterized by $\omega \log\left(\frac{P}{\hat{P}}\right)$. As $\hat{P}$ is the only choice variable, the numerator in the log can be ignored. The log of the determinant of a matrix is equivalent to a sum of the log of the eigenvalues, resulting in the expression defined above.

Mutual Information pertains to an estimate of $\alpha = 1$, or a linear relationship between attention and the cost-to-benefit ratio of acquiring that information, whereas Fisher information, as in Hébert and Woodford (2021), corresponds to a nonlinear relationship between the cost-to-benefit ratio of a dimension of information, $\alpha = \frac{1}{2}$.

In the Section 2, I propose both a framework and estimator to measure the shrinkage matrix $(\mathbb{I} - \Omega)$. The measures of attention can than be taken as moments to estimate the parametric values of ω and α for different types of forecasters — I estimate these parameters in Section 3.

2. A FACTOR ANALYSIS OF FORECASTS MADE UNDER RATIONAL INATTENTION

Forecasters optimally allocate attention along the same dimensions as the principal components (factors) taken from the implied time series panel of forecasts. Optimal attention allocation is inherently a lower-dimensional decision, as forecasters will choose to focus more on underlying economic processes that are important for more of the forecasted variables. Attention simply scales up or down the strength of each factor.

I take the optimal forecast from a rationally inattentive forecaster as the data generating process for a time series panel of forecasts and examine the properties of a singular value decomposition. I show that a time series forecast panel made by a rationally inattentive forecaster with separable spectral information costs can be expressed in terms of factors and loadings from a forecast made by a full information forecaster (0 cost of information acquisition) and a diagonal attention matrix. The elements of this matrix capture the forecaster's degree of attentiveness to a specific dimension of information (e.g. inflation, interest rates) and reflect the relative importance of the dimension of information, along with the forecaster's cost of information acquisition. Lastly, I propose an estimator for this shrinkage matrix.

2.1. *Attention as Shrinkage in Factor Decomposition*

Relative to the factors estimated from a "full information" forecast — a time series panel of predictions made by someone who has no cost of information acquisition — the factors estimated from the rationally inattentive forecast will be shrunken in magnitude. The weight will depend on the relative importance of the factors, along with the rationally inattentive forecasters implicit cost of information acquisition.

Forecasts are now made at each time t, and so include time subscripts. Full information forecasts are denoted with superscript $*$.

$$\underset{(n\times 1)}{X_t^*} = \underset{(r\times 1)}{B}\underset{(r\times 1)}{\mathbf{a}_t}$$

In each period, the rationally inattentive forecaster chooses their target posterior variance and acquires a signal with a precision to achieve that posterior variance. They then make their forecast according to their belief of that state:

$$\begin{aligned}\underset{(n\times 1)}{X_t} &= \underset{(n\times r)}{B'}\underset{(r\times 1)}{\hat{\mathbf{a}}_t} \\ &= B'\left(\mathbb{I} - U^B\Omega U^{B\prime}\right)\mathbf{a}_t\end{aligned}$$

With the singular value decomposition of B, $B = U^B D^B V^{B\prime}$ the attention matrix Ω is absorbed as a shrinkage matrix along the eigenvalues of B:

$$X_t = V^B D^B\left(\mathbb{I} - \Omega\right)U^B\mathbf{a}_t + B'\left(\mathbb{I} - U^B\Omega U^{B\prime}\right)\xi_t$$

$$\text{where } \xi_t = Z'\left(ZZ' + Q\right)^{-1}\epsilon_t$$

The fundamental shocks are normally distributed with unit varince: $\mathbf{a}_t \sim N\left(0, \mathbb{I}\right)$. There is no autocorrelation. As such, the last term ($U^B\mathbf{a}_t$) simply represents a rotation of the shocks from the standard basis ($\mathbb{I}$) to a new basis along the eigenvectors of B. These are then scaled by both the loadings that only reflect the B matrix, ($V^B D^B$), along with the forecaster's choice of optimal attention allocation (($\mathbb{I} - \Omega$)).

Moving into the matrix time series panel format, these shocks are simply extended along the time dimension — each of which is then captured as a principal component. Define the matrix of fundamental shocks as $\mathbf{A} = \left(\mathbf{a}_1', \mathbf{a}_2', \ldots \mathbf{a}_T'\right)'$.

$$\underset{(T\times N)}{X} = \underset{(T\times r)}{\mathbf{A}}\underset{(r\times r)}{U^B}\underset{(r\times r)}{D^B}\underset{(r\times r)}{\left(\mathbb{I} - \Omega\right)}\underset{(r\times N)}{V^{B\prime}} + \underset{(T\times r)}{\xi}\underset{(r\times N)}{B}$$

$$\text{where } \xi = \epsilon Z'\left(ZZ' + Q\right)^{-1}$$

Compare this to the same decomposition for a full information forecaster:

$$\underset{(T\times N)}{X^*} = \underset{(T\times r)}{\mathbf{A}}\underset{(r\times r)}{U^B}\underset{(r\times r)}{D^B}\underset{(r\times N)}{V^{B\prime}}$$

The latent shocks $\mathbf{A}$ are rotated from their original basis ($\mathbb{I}$) to the basis that characterizes sequentially descending most important dimensions for the vector of forecasted variables (U^B). The eigenvalue matrix D^B scales these dimensions to the degree of importance in the forecast. The rationally inattentive forecast matrix, relative to the full information forecast matrix, includes an additional matrix which shrinks and thresholds these singular values. The error in forecaster signals, ϵ^j, is first mapped to the dimensions of the latent shocks $\mathbf{A}$ - the signals need not be along the same basis as $\mathbf{A}$. These rotated errors then get mapped into the forecasted variables by B. The error in forecasts will also reflect the forecaster's cost of information acquisition.

Note that the expected variance of the shocks is unity, $\text{Var}(\mathbf{A}) = \mathbb{I}$. The term $\mathbf{A}U^B$ represents a new set of eigenvectors (the basis U^B extended along the time dimension). This representation makes it straightforward to observe the factors and loadings from the best possible forecast[9]:

$$\tilde{X}^* = F^*\Lambda^*$$
$$(F^*, \Lambda^*) = \left(\mathbf{A}U^B, D^B V^{B\prime}\right)$$

The number of common factors could be as many as the number of variables being forecasted $r \leq N$. If $r < N$, then the forecast will be completely determined by fewer common factors than variables. In the empirical application, the best possible forecast is a proxy, and thus has error. I use the information criteria 2 econometric procedure of Bai and Ng (2002) to estimate the number of common factors.

PROPOSITION 2: *If rationally inattentive forecasters and full information forecasters share the same weight (W), coefficient ($\tilde{B}$), and prior variance matrix (P), the forecast made by a rationally inattentive forecaster who faces Separable Spectral information costs can be written in terms of an attention matrix and the factors and loadings estimated from the full information forecast.*

$$\tilde{X} = F^* \left(\mathbb{I} - \Omega\right) \Lambda^{*\prime} + e$$

[9] For a review on estimation of principal components, readers are directed to Stock and Watson (2002a,b), Bai and Ng (2002, 2019).

The informationally-constrained common component includes an extra shrinkage and thresholding matrix $(\mathbb{I} - \Omega)$ as compared to the full information forecasts. This thresholded decomposition parallels the rank regularized estimation of approximate factor models, as in Bai and Ng (2019) and Cai et al. (2010). informationally-constrained agents are in effect performing singular value thresholding on the information space, whereas the full information agents are performing a singular value decomposition without thresholding. The effect is that rationally inattentive forecasts have shrunken and thresholded singular values as compared to the full information forecasts. The decisions by agents to ignore dimensions of the information space for which the cost does not exceed the benefit is reflected in the dimensionality of their forecast — their forecasts will have lower dimensions. Additionally, the amount that agents learn about dimensions of the information space are present in the degree of shrinkage (determined by the cost parameter, ω) that agents apply.

2.1.1. *Extension: Optimal Attention Allocation with Dynamic Factors*

The dynamic problem has distinct predictions from the static problem, as information will have more value for more persistent processes.[10] In many cases, however, forecasters receive information on the accuracy of their forecasts that negates any continuation value of signals received. For example, someone making forecasts of the 3rd quarter of U.S. GDP in August might track surprise innovations in retail spending or new orders in order to gain signals about consumer spending and investment. Following the release of the advance estimate of 3rd quarter GDP in late October, the informational content of the retail sales signals from August would be subsumed by the GDP release. The signal on consumer spending from the Bureau of Economic Analysis data has far higher precision than the average forecast. In this case, there would be no continuation value of information. In my analysis of a dynamic factor model, I include an additional assumption to capture this.

ASSUMPTION 2: *After making a forecast for variables at t and before making a forecast for variables at $t+1$, all forecasters receive a perfect signal of the state $\mathbf{a}_t$.*

[10] Afrouzi and Yang (2021) explore the resulting changes in attention allocation in the case with dynamic factors.

In the dynamic case, the agent's belief about the state, $\hat{\mathbf{a}}_t$, reflects the known persistence, characterized by the coefficients of $\mathcal{T}$:

$$\hat{\mathbf{a}}_t = \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k} + \left(\mathbb{I} - U^B \Omega U^{B\prime}\right)\left(\mathbf{a}_t - \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k}\right) + \xi_t$$

$$\text{where } \xi_t = Z'\left(ZZ' + Q\right)^{-1}\epsilon_t$$

Under dynamic factors, forecaster attention is allocated towards innovations in the underlying state, not the autoregressive component of the state process.

ASSUMPTION 3: $\left(\mathbf{a}_t - \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k}\right) \sim N\left(0, \mathbb{I}\right)$

The rationally inattentive forecasters problem is similar to that of a Nowcaster (Giannone et al. (2008)). Agents acquire signals on monthly data releases to update their beliefs of the underlying monthly state variables, which then get mapped into optimal forecasts through a bridge equation. The main difference between the Nowcaster and a rationally inattentive forecaster, however, is that the posterior variance, through a selection of their observation equation error variance (Q), is an endogenous choice. Agents choose how precise their signals are, which then determines the strength of the mapping from signal surprise to forecast. In the case of forecasting GDP, analysts could purchase information beyond what is publicly released by government statistical agencies. They could look at or purchase higher-frequency data to get a better signal on underlying economic growth. These additional signals come at an additional cost to the analyst.

The intuition for how attention is mapped into the factor structure of forecasts from the static case carries over to dynamic factors. The rationally inattentive forecast becomes:

$$\hat{X}_t = \underbrace{B' \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k}}_{\text{Autoregressive}} + \Lambda^* \underbrace{\left(\mathbb{I} - \Omega\right)}_{\text{Attention Matrix}} F^* + B'\xi_t$$

The full information loadings and factors are:

$$\left(\Lambda^*, F^*\right) = \left(V^B D^B, U^{B\prime}\left(\mathbf{a}_t - \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k}\right)\right)$$

In the dynamic case, the common factors are now innovations to the state rotated to the B matrix eigenvector basis. The matrix $(\mathbb{I} - \Omega)$ again captures the attention of a rationally inattentive agent to a dimension of the information space relative to the full information agent. Compare this to a full information agent's forecast, in which the forecast does not include the extra attention matrix:

$$X_t^* = B' \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k} + \Lambda^* F^*$$

In the empirical analysis that follows, all forecast data is orthogonalized to previous lags of the realized variables, thus isolating the component affected by attention.

2.2. *An Estimator of Forecaster Attention*

The previous section provides a theoretical framework that maps forecaster attention to the factor decomposition of forecasts. Given an empirical proxy for a full information forecast, the elements of the attention matrix can be estimated. In this section, I propose such an estimator and provide a theoretical justification for using the Consensus (median) forecast as the full information proxy.

My proposed estimator for attention estimates the elements for the diagonal and off-diagonal elements of the attention matrix, which will permit an analysis of the degree of similarity between forecast models. The rationally inattentive forecast with generalized attention matrix Φ can be written as:

$$\hat{X} = F^* \Phi \Lambda^{*\prime} + e$$

The matrix of factors can be written as a concatenation of vectors of the individual factors: $F_t^* = [F_{1t}^*, F_{2t}^*, \ldots, F_{rt}^*]$. The $N \times r$ loadings matrix Λ^* can be written as $\Lambda^* = (\lambda_1, \lambda_2 \ldots \lambda_N)'$. The vector of loadings corresponding to variable i can further be written as $\lambda_i^* = [\lambda_{1i}^*, \lambda_{2i}^*, \ldots, \lambda_{ri}^*]$. The i^{th} forecasted variable by the rationally inattentive forecaster at time t can be written as:

$$\begin{aligned}
\hat{X}_{i,t} = & F_{1t}^* \left(\lambda_{1i}^* \Phi_{1,1} + \lambda_{2i}^* \Phi_{1,2} + \cdots + \lambda_{ri}^* \Phi_{1,r}\right) \\
& + F_{2t}^* \left(\lambda_{1i}^* \Phi_{2,1} + \lambda_{2i}^* \Phi_{2,2} + \cdots + \lambda_{ri}^* \Phi_{2,r}\right) \\
& + \ldots
\end{aligned}$$

$$+ F^*_{rt}\left(\lambda^*_{1i}\Phi_{r,1} + \lambda^*_{2i}\Phi_{r,2} + \cdots + \lambda^*_{ri}\Phi_{r,r}\right) + e_{i,t}$$

ASSUMPTION 4: *For factor* $k \in 1, \ldots, r$

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\lambda^*_i e_{it} F^*_{k,t}\right) \xrightarrow{d} N(0, \Sigma_{\Lambda e F_k})$$

where

$$\Sigma_{\Lambda e F_k} = \underset{N\to\infty, T\to\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\lambda_i\lambda_j'\mathbb{E}\left[e_{it}e_{js}F_{k,t}F_{k,s}\right]$$

Given estimates of factors and loadings from the full information proxy, $\left(\tilde{F}^*, \tilde{\Lambda}^*\right)$. The proposed estimator of the rotated attention matrix proceeds in two steps:

1. For each $i \in 1, \ldots, N$ regress $T \times 1$ vector $\hat{X}_i$ on $T \times r$ matrix $\tilde{F}^*$. This results in N estimates for the k^{th} $(k \in 1, \ldots r)$ coefficient: $\left(\lambda^*_{1i}\Phi_{k,1} + \lambda^*_{2i}\widehat{\Phi_{k,2} + \cdots + \lambda^*_{ri}\Phi_{k,r}}\right)$.
2. For each $k \in 1, \ldots, r$ regress N coefficients corresponding to k^{th} factor from first step onto $N \times r$ matrix $\tilde{\Lambda}^*$. The resulting vector of coefficients are estimates of the k^{th} row of Φ, $(\widehat{\Phi_{k,1}}, \widehat{\Phi_{k,2}}, \ldots \widehat{\Phi_{k,r}})$.

The asymptotic distribution of the estimator is summarized in Proposition 3.

PROPOSITION 3: *Suppose Assumption 4 holds. Under the normalization* $\frac{F'F}{T} = \mathbb{I}$*, the proposed estimator for the* k^{th} *row* $(k \in 1, \ldots, r)$ *of the rotated attention matrix* Φ *converges to the following distribution.*

$$\sqrt{TN}\left(\widehat{\widehat{(\Phi)}'}_{k,1:r} - \Phi'_{k,1:r}\right) \to_d N\left(0, \Sigma_\Lambda^{-1}\Sigma_{\Lambda e F_k}\Sigma_\Lambda^{-1}\right)$$

Where $\Sigma_{\Lambda e F_k}$ *is given in Assumption 4, and* $\Sigma_\Lambda = \frac{\Lambda'\Lambda}{N}$.

More details on the derivation of the asymptotic distribution, along with Monte Carlo simulations can be found in the Appendix.

2.3. *Eigenvectors of Consensus and Best Possible Forecast*

The Consensus forecast (median forecasted value for each variable) is an ideal proxy for a full information forecast. The measure of attention captures both the shrinkage along

dimensions of the information space ranked in order of importance, along with rotation away from that ideal basis. If forecasters have independent and normally distributed forecast errors, then the eigenvector basis of the Consensus forecast will be equivalent to the theoretical full information forecast.

Both $\mathbf{a}$ and ϵ_{jt} are normally distributed and independent across forecaster, thus the distribution of X_{it}^{j} will also be normally distributed. Under a normal distribution, the median and mean will be the same and equal to the expectation of the forecast over forecaster type, $\mathbb{E}[\cdot]$. The best possible $(*)$, j^{th} percentile (j), and Consensus (C) forecasts can be written as:

$$X^{*} = \mathbf{A}B$$

$$X^{j} = \hat{\mathbf{A}}_{j}B$$

$$X^{C} = \mathbb{E}_{j}\left[\hat{\mathbf{A}}_{j}B\right]$$

The matrices $\mathbf{A}$ and $\hat{\mathbf{A}}_{j}$ are both of dimension $T \times r$. The singular value decomposition of X^{*} is the same as in Section 3.

$$X^{*} = \mathbf{A}U^{B}D^{B}V^{B\prime}$$

Let $\xi^{j} = \epsilon^{j}Z_{j}'\left(Z_{j}Z_{j}' + H_{j}\right)^{-1}$. Compared to the full information forecast, the singular value decomposition for the j^{th} percentile forecaster includes an extra term in the time dimension being rotated by the U^{B} eigenvectors ($\xi^{j}U^{B}D^{B}\left(\mathbb{I} - \Omega_{j}\right)U^{B\prime}D^{B-1}$), as well as the scaling attention matrix on the eigenvalue matrix.

$$\begin{aligned} X^{j} &= \mathbf{A}U^{B}\left(\mathbb{I} - \Omega_{j}\right)D^{B}V^{B\prime} + \xi^{j}B \\ &= \left(\mathbf{A} + \xi^{j}U^{B}D^{B}\left(\mathbb{I} - \Omega_{j}\right)U^{B\prime}D^{B-1}\right)U^{B}\left(\mathbb{I} - \Omega_{j}\right)D^{B}V^{B\prime} \end{aligned}$$

In the case of the Consensus forecast, only the eigenvalues are scaled by the average attention matrix, as the error term in expectation across forecaster is 0.

$$\begin{aligned} X^{C} &= \mathbf{E}_{j}\left[\mathbf{A}\mathbf{U}^{B}\left(\mathbb{I} - \Omega_{j}\right)D^{B}V^{B\prime} + \epsilon^{j}Z_{j}'\left(Z_{j}Z_{j}' + H_{j}\right)^{-1}B\right] \\ &= \mathbf{A}U^{B}\left(\mathbb{I} - \mathbf{E}_{j}\left[\Omega_{j}\right]\right)D^{B}V^{B\prime} \end{aligned}$$

In the applications considered, the goal is to measure attention, or the scale of importance of the relevant eigenvector in the information space, relative to the best possible forecaster. In general it is not possible to observe the best possible forecaster, but it is possible to observe the Consensus forecast. The Consensus forecast provides the appropriate eigenvector basis along which to measure attention.

3. EMPIRICAL FEATURES OF MEASURED FORECASTER ATTENTION

The main empirical finding in this Section is that, contrary to what is implied by a linear cost on Mutual Information, forecaster attention in general displays a nonlinear relationship with the cost-to-benefit ratio of a dimension if information. Shifts in the value of more important sources of information will drive larger moves in attention than shifts in the value of less important sources of information. Before illustrating this point using forecaster attention to information associated with long-term inflation, I present the empirical estimates of forecaster attention matrices — both the main diagonal elements and the off-diagonal elements. The results show that all forecasters, top-, middle- and bottom-performing, share the same reduced-form forecast model. Forecasters differ in their degree of attentiveness to relevant information.

I use forecast data from the *Survey of Professional Forecasters* (SPF). The SPF is an anonymous quarterly survey of professional macroeconomic forecasters. I focus on data from 1985 until 2019 ($T = 140$) for forecasts made for a large number of macroeconomic variables for multiple forecast horizons.[11] Individuals surveyed by the SPF have an identifier associated with them, but drop in and out of the sample and do not stay in the sample for the duration of the period under consideration. Following a similar method used by Giaco-

[11] The following set of variables: nominal GDP, real GDP, real Consumption, business fixed investment, residential investment, inventories, corporate profits, industrial production, employment (nonfarm payroll), unemployment, housing starts, GDP price index, CPI, real net exports, real government spending (federal and state and local), Treasury Bills (3 Month), Treasury Bonds (10 Year), Treasury Spreads (10 Year - 3 Month), and BAA corporate bond spreads over treasuries. Forecasts of the percent change, change (interest rates, exports, inventories, employment), or level (Treasury and BAA spreads) are used for each series. Forecasts are taken at 6 different horizons: current quarter ($t+0$), subsequent four quarter ($t+h,\ h=1,2,3,4$), and annual forecasts for two calendar years ahead ($t+\overline{2Y}$). Forecasts for percent changes and changes are for the annual average two calendar years ahead relative to the annual average one calendar year ahead. If a forecast is made in 2004, this will correspond to the forecast of 2006 relative to 2005, regardless of when in 2004 the forecast is made.

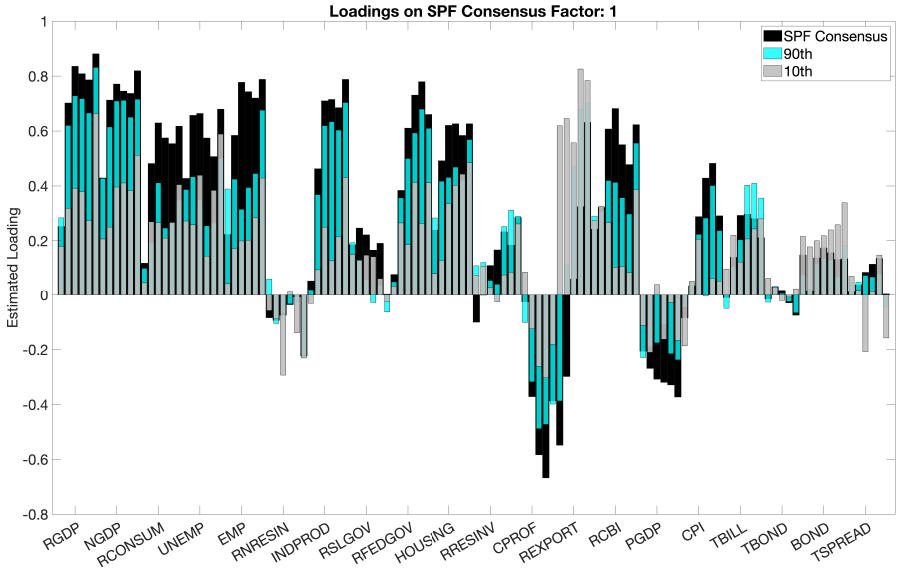


FIGURE 1.—Figure shows the estimated coefficients, $\hat{\beta}^{j,C}_{v,h,1}$, from regressing the forecast made by forecaster type $j = 10, 90$ of variable v at the h quarter horizon, $X^{j}_{v,t+h}$, on the first principal component of the Consensus forecast (C). Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon. $X^{j}_{v,t+h} = \hat{\beta}^{j,C}_{v,h,1}\tilde{F}^{C}_{1,t} + u^{j,C}_{v,t+h}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

letti et al. (2021) and Bianchi et al. (2022), I build a time series of forecasts for forecasters corresponding to different percentile types based on forecast performance. Forecasts are orthogonalized to the lagged real-time realizations of the underlying variables to remove autocorrelation. More details on the construction of forecasts, along with evidence of the persistence of forecaster rank, can be found in the Appendix.

3.1. *Measured Forecaster Attention*

The measure of forecaster attention captures the average relative magnitude loadings on factors between a given forecaster and the proxy for full-information forecaster. The loadings are on factors taken from the full information forecaster, but reflect the importance dimensions of information.

Forecaster attention is captured by the average ratio forecaster and Consensus loadings on a common factor. In Figure 1, I plot the loadings on the first principal component taken

from the Consensus forecast panel. In Figure 2, I show the average ratio of loadings between forecaster and Consensus for the the first eight principal components. The attention along the first factor, as captured in Figure 1, is captured by the first bar in each subplot in Figure 2. Relative to a Consensus benchmark, the 10th percentile forecasters pay 50% attention whereas the 90th percentile forecasters display closer to 80% attention, as captured by the average relative magnitude of factor loadings.

In Figure 2, I present measured attention to the first eight Consensus principal components. Attention is allocated along principal component dimensions, or the dimensions that explain the most variance in descending order. The average ratio of loadings in Figure 1 are captured by the first bar in the 10th and 90th percentile forecaster subplots. The factors are not restricted to be aligned with specific economically-meaningful dimensions, but correlations between the factors and the forecasted variables provide some intuition of what the dimensions represent. In the Appendix, I present the top variables by correlation with each principal component. The figure reveals a specific pattern in attention across principal components. All forecasters display less attention to less important sources of information. Forecasters pay the most attention to the first principal component, which is correlated with measures of real activity. The second, third, and fourth principal components are correlated with unemployment and employment variables, inflation variables, and housing-related variables, respectively. Forecasters in general pay less attention to the "inflation factor" (third bar in each subplot), than to the "real activity" factor (first bar). Figure 3 presents forecaster attention relative to the Federal Reserve Greenbook forecast. The pattern in attention is similar.

The forecasters downweight the less important sources of information, and in some cases completely ignore the information. Estimates of the number of common components for each type of forecaster provide evidence of this potential thresholding. Information criteria tests estimated 20 common factors in the Consensus forecast, and only 8, 7, and 7 common factors for the 90th, 50th and 10th percentile forecasters. Forecasters have a lower-dimensional forecast structure than the Consensus forecast.

3.1.1. *Evidence of Similarities in Reduced-Form Forecast Models*

The estimated factors from each type of forecast are correlated with similar groupings of variables, but the estimator for attention permits a more robust analysis of differences

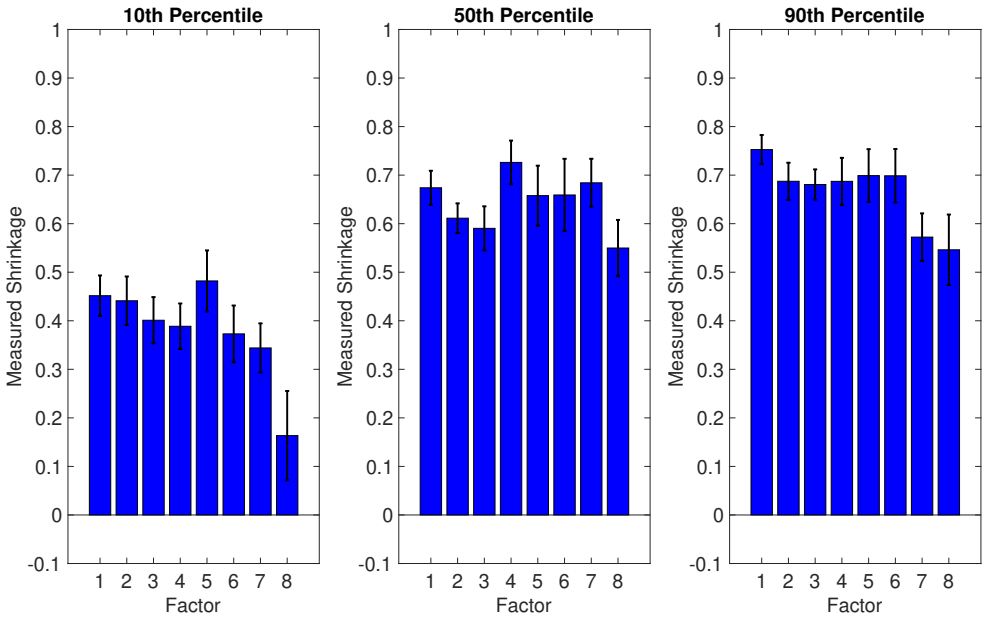


FIGURE 2.—Attention relative to *Survey of Professional Forecasters* Consensus forecast. Figure shows the estimated diagonal elements of the attention matrix, $\left(\mathbb{I} - \Omega_{j,i}\right)$ for $i = 1, \ldots 8$ and $j = 10, 50, 90$.

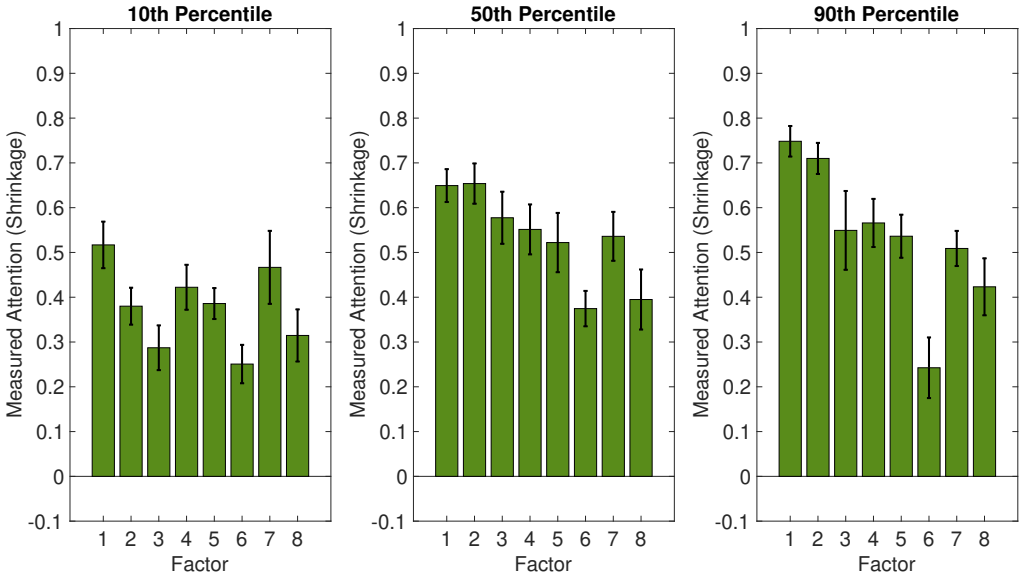


FIGURE 3.—Attention relative to the Federal Reserve Greenbook Forecast. Figure shows the estimated diagonal elements of the attention matrix, $\left(\mathbb{I} - \Omega_{j,i}\right)$ for $i = 1, \ldots 8$ and $j = 10, 50, 90$.

between forecaster models. In particular, if forecasters had differences in reduced form models, there would be no reason for the attention matrix to be diagonal. I find, however, that the estimated attention matrix (relative to a Consensus benchmark) of forecasters is diagonal, implying that forecasters share a reduced-form forecast model, and instead differ on their degree of attentiveness.

The reduced form forecast model of a full information forecaster is characterized by the matrix B, which map innovations in the underlying shocks $\mathbf{a}$ into an optimal forecast. If a forecaster had a different model than the full information forecaster, say $C \neq B$, then the attention matrix would no longer be diagonal. The attention matrix would instead be rotated by matrices H and $\tilde{H}$ reflecting eigenvectors and values of the C matrix. This would in general lead to a non-diagonal attention matrix.

$$\hat{X} = F^{*} H \left(\mathbb{I} - \Omega\right) \tilde{H}' \Lambda^{*\prime} + e$$
$$H = U^{B} U^{C\prime}$$
$$\tilde{H}' = D^{C} V^{C} V^{B\prime} \left(D^{B}\right)^{-1}$$

The estimates of the full attention matrix, Φ, for the 50th percentile forecaster is presented in Figure 4. The matrix is mostly diagonal indicating that there is essentially no rotation away from the most important dimensions of information.[12] Estimated attention matrices for the 10th and 90th percentile forecasters are similarly mostly diagonal. The finding of a mostly diagonal attention matrix is consistent with the empirical features of principal components estimated from each type of forecast.

3.2. *Estimates of Cost Function Parameters*

How individuals allocate attention has important macroeconomic consequences, such as driving nonfundamental volatility in economic growth. Equipped with the theoretical mapping from attention governed by Separable Spectral information costs to forecast factor structure and the estimator for measuring attention, I estimate how attention varies with information importance and cost. A cost function characterized by mutual information is in

[12] In comparison, attention relative to a "full information" forecast generated using machine learning methods from a large macroeconomic dataset does not display a mostly diagonal pattern.

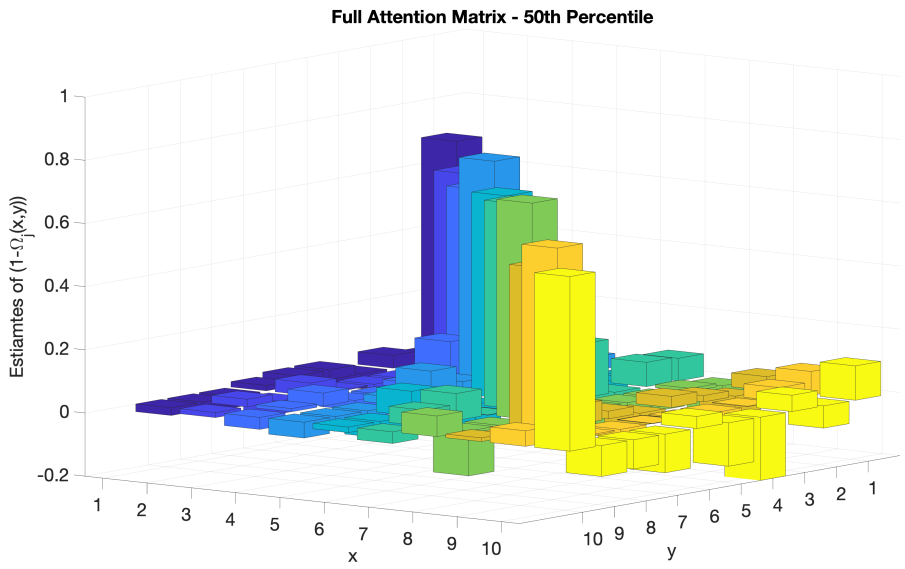


FIGURE 4.—Point estimates of full attention matrix $(\mathbb{I} - \Omega_{50})$, for the 50th percentile forecaster.

general rejected — forecasters display a nonlinear relationship between attention allocation and the cost-to-benefit ratio of information.

As described previously, the optimal attention of a forecaster of rank j with information cost ω_j, and an optimal forecast characterized by the matrix B, and a general parametric form of Separable Spectral information costs to dimension i is characterized by:

$$1 - \Omega_{j,i} = 1 - \left(\frac{\omega_j}{\lambda_i^B}\right)^{\alpha_j}$$

A parameter value of $\alpha = 1$ corresponds to attention under a cost on mutual information, whereas a value of $\alpha = 0.5$ corresponds to attention under Fisher information. Parameters ω_j and α_j are fixed across information dimensions, providing sufficient moments for estimation. I jointly estimate the information cost ω_j and curvature α_j parameters for forecasters by performance rank using minimum distance. Estimation details are in the Appendix.

The estimated curvature parameters are documented in the right panel of Figure 5, along with the empirically estimated attention curves. The left panel shows theoretical attention curves for a specific form of cost function. As forecast performance declines the nonlinearity of attention increases. Estimates of the information cost parameter by forecast per-

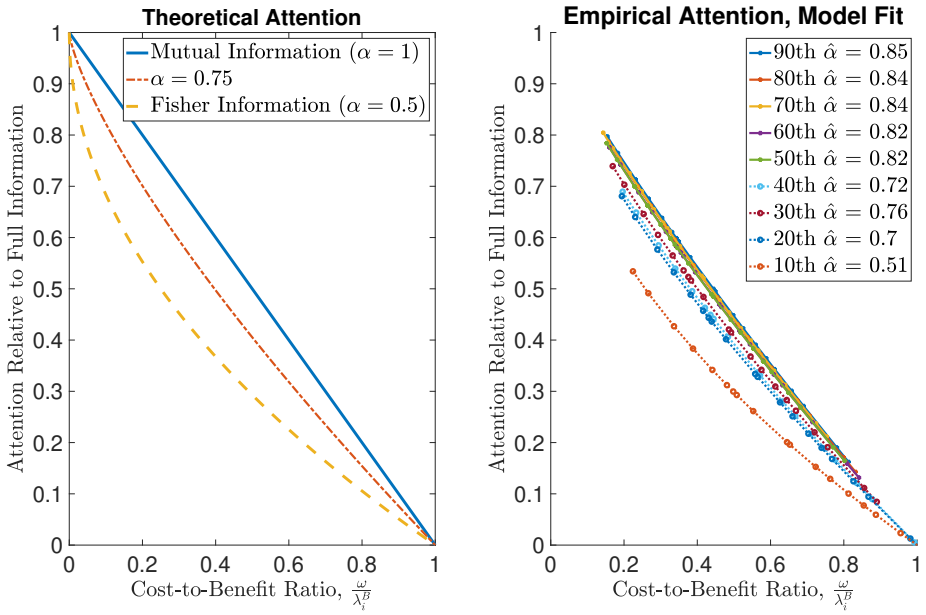


FIGURE 5.—Model-implied attention for cost-to-benefit ratio $\frac{\lambda^B}{\hat{\omega}_j}$ for j^{th} percentile forecasters theoretical (left) and estimated (right). The Model Fit attention is given by $1 - \left(\frac{\hat{\omega}_j}{\lambda_i^B}\right)^{\hat{\alpha}_j}$.

formance, ω_j, are documented in the Appendix. Better-performing forecasters have lower information costs.

3.3. *Evidence of Nonlinearity in Forecaster Attention to Long-Term Inflation Factor*

To illustrate nonlinearities in attention, I look at forecaster attention to long-term inflation under different regimes. The importance of a dimension of information reflects both the weight that a factor receives in a forecast (how importance is the dimension of information to the overall forecast), along with the forecaster's prior belief about the variance of a factor. If a forecaster has a higher prior uncertainty about long-term inflation they will optimally allocate more attention toward it.

3.3.1. *Attention to Economically-Meaningful Factors*

If the common forecast model has a specific structure, then the factors can be rotated to correspond to economically-meaningful dimensions. Following the logic of the identification restriction PC2 in Bai and Ng (2013), if there is a dependence structure such that forecasted variables sequentially capture variation in the other variables, a valid identifica-

tion scheme exists. Consider an agent forecasting macroeconomic variables. The forecast of underlying growth would be captured in the forecast of GDP. Underlying economic growth would also be reflected in the forecast for inflation, along with an additional inflation component. Econometrically this would be captured by the top part of the coefficient matrix B being upper triangular with nonzero entries along the diagonal.[13]

PROPOSITION 4: *If the coefficient matrix B has the following structure, the first factor can then be associated with the forecasted variable ordered first, the second factor with the second ordered variable (orthogonal to the first factor), and so on:*

$$B' = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, B_1 = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ b_{21} & b_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \dots & b_{pp} \end{pmatrix}$$

$$b_{ii} \neq 0,\ i = 1, 2, \dots, p$$

Furthermore, identified factors will have unit variance and loadings of the best possible forecast will have the following structure:

$$\Lambda^* = \begin{pmatrix} \Lambda_1 \\ \Lambda_2 \end{pmatrix}, \Lambda_1 = \begin{pmatrix} \lambda_{11} & 0 & \dots & 0 \\ \lambda_{21} & \lambda_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \dots & \lambda_{pp} \end{pmatrix}$$

$$\lambda_{ii} \neq 0,\ i = 1, 2, \dots, p$$

Estimating these factors requires an additional step after taking principal components. Let $\hat{F}^*$ and $\hat{\Lambda}^*$ denote factors and loadings estimated via principal components with the forecasted variables economically ordered. Given a QR decomposition of $\hat{\Lambda}_1^{*\prime}$ to yield $\hat{\Lambda}_1^{*\prime} = Q \cdot R$ with R an upper triangular with positive diagonal elements and Q an orthogonal

[13] This restriction subsumes the identification restriction of the top part of B being diagonal, which would indicate a set of the forecasted variables acting as an already orthogonal basis for the rest of the forecasted variables

matrix such that $Q'Q = \mathbb{I}$, the rotated and identified factors ($\tilde{F}^*$) and loadings ($\tilde{\Lambda}^*$) are as follows:

$$\tilde{F}^* = \hat{F} \cdot Q \qquad\qquad \tilde{\Lambda}^* = \hat{\Lambda}^* \cdot Q = \left[R', \hat{\Lambda}_2^{*\prime}\right]'$$

Attention can then be measured along the new basis of factors. If the forecaster's economic model satisfies the assumption of Proposition 4 (part of the matrix mapping shocks to loadings upper triangular), then attention can be measured along economically meaningful dimensions. Factors and loadings from the full information benchmark are estimated via standard principal components are rotated to align with the new indicated basis.

When the underlying state process has persistence, information associated with near term shocks will be more valuable for the forecaster as this information should affect long-term forecasts through the persistence channel. Consider the case in which $q = 1$; the VAR process of the states is of order 1. Agents make forecasts of a vector of variables, X_t. This vector can include variables that are realized in the current quarter (akin to Nowcasting) or can be forecasts of variables that will be realized further out. The vector X_t could include forecasts of GDP growth for this quarter, forecasts of GDP growth for next quarter, and forecasts of GDP growth two years out. In the empirical exercise, forecasts of variables at multiple horizons are included.

Let X_t be composed of forecasts at short (S) and long (L) horizons for two different variables: GDP growth ($x^g_{S,t}$, $x^g_{L,t}$) and inflation ($x^\pi_{S,t}$, $x^\pi_{L,t}$). There are 4 latent factors which correspond to: short-term growth, short-term inflation orthogonal to growth, long-term growth orthogonal to short-term factors, and long-term inflation orthogonal to everything else.

$$X_t = \begin{pmatrix} x^g_{S,t} \\ x^\pi_{S,t} \\ x^g_{L,t} \\ x^\pi_{L,t} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ b_{31}\,(\rho_{g,s}) & 0 & b_{33} & 0 \\ b_{41}\,(\rho_{g,s}) & b_{42}\,(\rho_{\pi,S}) & b_{43} & b_{44} \end{pmatrix}$$

$$\mathbf{a}_t = \begin{pmatrix} \mathbf{a}^g_{S,t} \\ \mathbf{a}^\pi_{S,t} \\ \mathbf{a}^g_{L,t} \\ \mathbf{a}^\pi_{L,t} \end{pmatrix} \qquad \mathcal{T} = \begin{pmatrix} \rho_{g,S} & 0 & 0 & 0 \\ 0 & \rho_{\pi,S} & 0 & 0 \\ 0 & 0 & \rho_{g,L} & 0 \\ 0 & 0 & 0 & \rho_{\pi,L} \end{pmatrix}$$

Short-term inflation load both on the short-term growth and short-term inflation-specific factor. Long-term growth mechanically loads on the short-term growth factor (through the autoregressive coefficient in the $\mathcal{T}$ matrix), as well as on a long-term growth specific factor. Long-term inflation loads on all factors.

The first factor, short-term growth, affects all of the forecasted variables. It affects the long-term forecasts through the persistence of the factor (captured by $\rho_{g/h,S}$). Intuitively, there are shocks that occur today that we know will affect some of these factors in the longer-term. Extreme changes in the commodities could eventually affect inflation, but the higher prices would take time to move through the production process. Given the persistence of the factors, innovations to the near-term factor will directly affect long-term forecasts. If the coefficients are high enough there will be an incentive to pay more attention to the short-term factors.

Returning to data from the *Survey of Professional Forecasters*, I assume that forecasts are driven by 4 latent factors, which are aligned with near-term real GDP growth (one quarter ahead), near-term CPI growth outside of near-term real GDP growth, long-term GDP growth (2 year ahead annual forecast) outside near-term real GDP and CPI growth, and long-term CPI growth outside of near-term real GDP and CPI and long-term GDP. In Figure 6, I show measured attention to a factor aligned with long-term inflation, outside of near-term growth and inflation and long-term growth, in three different inflationary regimes. In particular I consider a low-inflation regime (2010-2019, less than 2% average CPI inflation) after the Global Financial Crisis and before the post-Covid inflationary period, a medium inflation regime (1985-2007, around 3% average CPI inflation), and a high inflationary regime (2020-2023, close to 5% average inflation). Attention varies the most for the top-performing, or 90th percentile forecaster. Attention between the medium and low inflation regimes are essentially unchanged for the poorly- and medium-performing forecasters. Only the top-performing forecasters display attention that varies between the low- and medium-inflation regimes.

4. CONCLUSION

In this paper, I present a framework to measure forecaster attention to different dimensions of information. I show evidence that forecasters share similar reduced form forecast models, and instead differ on their degree of attentiveness to different sources of infor-

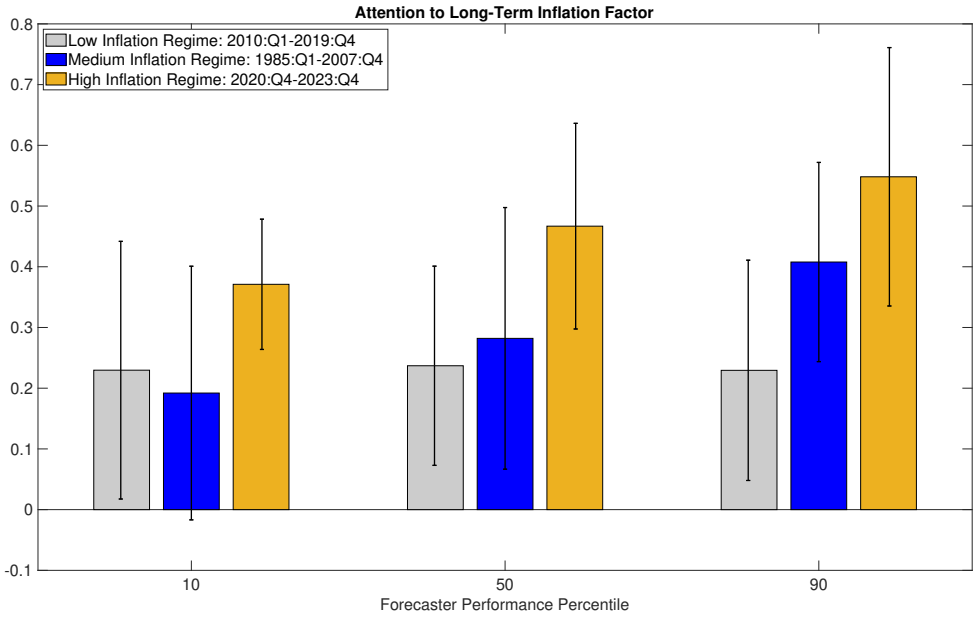


FIGURE 6.—Measured attention by forecaster type to information associated with long-term inflation (orthogonal to information associated with near-term growth, near-term inflation, and long-term growth), under low- (2010Q1-2019Q4, left bar), medium- (1985Q1-2007Q4, middle bar), and high- (2020Q4-2023Q4, right bar) inflationary regimes by forecaster performance percentile.

mation. Forecasters display less attention to less important sources of information, and in general display attention the varies nonlinearly with the cost-to-benefit of acquiring information. The empirical results suggest that researchers explore the use of cost functions beyond mutual information, such as those characterized by Fisher information. Separable spectral information costs provide a tractable alternative for the standard linear quadratic tracking problem.

APPENDIX A: SEPARABLE SPECTRAL COST FUNCTIONS

Similar to Caplin, Dean and Leahy (2022) and Hébert and Woodford (2021), I define a general information structure. Consider a $\mathbf{R}^d$ space. An information structure is specified by a collection of a prior, μ, measure over $\mathbf{R}^d$, a collection of posteriors, Γ (such that each $\gamma \in \Gamma$ is a measure over $\mathbf{R}^d$), and a measure $\mathcal{Q}$ over Γ which describes the unconditional probability of reaching a particular posterior. Given a scalar-valued function $H(\cdot)$ defined over possible measures on $\mathbf{R}^d$, the cost of information under this structure can then be written as:

$$C(\mu; \Gamma, \mathcal{Q}) = E_{\mathcal{Q}}[H(\gamma)] - H(\mu)$$

Cost functions require a specific definition of the function H. Separable spectral cost functions are defined over the variance implied by the probability measures. Before defining separable spectral costs formally, I introduce the concept of matrix convexity. A matrix-valued function is matrix-convex if for any scalar $0 \leq \alpha \leq 1$, and positive definite symmetric matrices A and B the following holds:

$$\mathbf{f}(\alpha A + (1-\alpha) B) \preceq \alpha \mathbf{f}(A) + (1-\alpha) \mathbf{f}(B)$$

Where $\preceq$ defines a *Löwner* order on symmetric $d \times d$ matrices:

$$A \preceq B \iff B - A \text{ is positive semidefinite}$$

DEFINITION 2: *Consider an information cost structure over a collection of Gaussian priors μ and posteriors Γ, and a distribution of posteriors $\mathcal{Q}$:*

$$C(\mu; \Gamma, \mathcal{Q}) = E_{\mathcal{Q}}[H(\gamma)] - H(\mu)$$

*A **separable spectral cost function** is characterized by the function H:*

$$H(\gamma) = tr[\mathbf{f}(\Sigma(\gamma))]$$

The matrix $\Sigma(\gamma)$ is the $d \times d$ variance implied by the measure γ. $\mathbf{f}$ is a spectral matrix-valued function defined on the set of symmetric positive definite $d \times d$ matrices, here denoted by A, matrices with eigenvalue decomposition $A = \sum_{i=1}^{d} \lambda_i u_i u_i'$. $\mathbf{f}$ can be written

as:

$$\mathbf{f}(A) = \sum_{i=1}^{d} f(\lambda_i) u_i u_i'$$

The function $f(\lambda)$ is a real-valued defined for all $\lambda > 0$ and is strictly monotonically decreasing with a positive range. The function $\mathbf{f}(A)$ is matrix-convex.

Both the specific posterior (γ) and the collection of posteriors Γ are restricted to Gaussian distributions. This comes from the fact that the states follow a Gaussian process and signals have Gaussian shocks. The forecaster, after observing the signal updates their belief about the realization of the shock from a mean zero, unit variance prior to a mean $Z(ZZ' + H)^{-1} s$, variance $\mathbb{I} - Z'(ZZ' + H)^{-1} Z$ posterior belief. Given the forecaster's choice of signal structure (Z and H), the only difference in Bayes-consistent posteriors is in the mean.

Separable spectral costs can be written as differences in the sums of the functions of eigenvalues of the prior (P) and posterior ($\hat{P}$) variance matrices, as the trace is the sum of the eigenvalues. Denote the eigenvalues by λ_i^0 and λ_i for $i \in 1, \ldots d$, respectively.

$$C^{SS}\left(P, \hat{P}\right) = \sum_{i=1}^{d} f(\lambda_i) - \sum_{i=1}^{d} f\left(\lambda_i^0\right)$$

A.1. *Posterior Beliefs and Variance Given Signals*

Forecasters get to choose signal structure Z and signal error variance Q, and receive signals according to:

$$\underbrace{s}_{m\times 1} = \underbrace{Z}_{m\times r}\underbrace{\mathbf{a}}_{r\times 1} + \underbrace{\epsilon}_{m\times 1} \qquad\qquad \epsilon \sim N(0, Q)$$

Forecasters have prior knowledge that $a \sim N(0, \mathbb{I})$. Following the standard Kalman filtering equations (Durbin and Koopman (2012), Chapter 4), the forecaster's Kalman gain is:

$$K = Z\left(ZZ' + Q\right)^{-1}$$

After receiving the signal, the forecaster has the following belief about the state $\hat{\mathbf{a}}$ and posterior variance $\hat{P}$:

$$\hat{\mathbf{a}} = K\left(Z\mathbf{a} + \epsilon\right)$$

$$\hat{P} = \mathbb{I} - KZ$$

Combining the two expressions yields:

$$\hat{\mathbf{a}} = \left(\mathbb{I} - \hat{P}\right)\mathbf{a} + K\epsilon$$

A.2. *Proof of Proposition 1*

The decision problem for the forecaster can be written as:

$$\min_{\hat{P}} \mathbb{E}\left[\text{tr}\left(BB'\hat{P}\right) + \omega \mathbf{f}(\hat{P})\right]$$

$$\text{s.t. } \mathbb{I} - \hat{P} \succeq 0$$

Using the decomposition $\hat{P} = UDU'$, d_i denotes the i^{th} diagonal element of D, and u_i the i^{th} column of U, the agent's choice can be written as:

$$\min_{\{d_i, u_i\}_i^r} \sum_{i=1}^{r} d_i u_i' BB' u_i + \omega \sum_{i=1}^{r} f\left(d_i\right)$$

$$\text{s.t. } 0 \leq d_i \leq 1$$

By the Courant-Fischer-Weyl min-max principle, the expression $u_i' BB' u_i$ is minimized at the eigenvalues and eigenvectors of BB' in descending order of the eigenvalues. For $i = 1$, the minimum occurss at the first eigenvector and eigenvalue of BB', u_1^B and λ_i^B, respectively. For $i = 2$, the minimum occurss at the second eigenvector and eigenvalue of BB', and so on. Hence for each $i \in 1, \ldots, n$, the minimization problem can be characterized as:

$$\min_{\{d_i\}} d_i \lambda_i^B + \omega f\left(d_i\right)$$

$$\text{s.t. } 0 \leq d_i \leq 1$$

The value of d_i which determines the minimium is characterized by $-\frac{\lambda_i^B}{\omega} = f'\left(d_i\right)$.

APPENDIX B: ASYMPTOTIC DISTRIBUTION OF RELATIVE ATTENTION ESTIMATOR

As described in section 4.1, Φ characterizes the rotated attention matrix, where $\Phi_{x,y}$ denotes the x^{th} row and y^{th} column. Let $\Phi = H\left(\mathbb{I} - \Omega\right)\tilde{H}'$. The rationally inattentive forecast can be be written as:

$$\hat{X} = F^{*}\Phi\Lambda^{*\prime} + e$$

The matrix of factors can be written as a concatenation of vectors of the individual factors: $F_t^* = [F_{1t}^*, F_{2t}^*, \ldots, F_{rt}^*]$. The vector of loadings corresponding to variable i can be written as $\lambda_i^* = [\lambda_{1i}^*, \lambda_{2i}^*, \ldots, \lambda_{ri}^*]$. The i^{th} forecasted variable by the rationally inattentive forecaster at time t can be written as:

$$\begin{aligned}\hat{X}_{i,t} = F_{1t}^{*}\left(\lambda_{1i}^{*}\Phi_{1,1} + \lambda_{2i}^{*}\Phi_{1,2} + \cdots + \lambda_{ri}^{*}\Phi_{1,r}\right) + F_{2t}^{*}\left(\lambda_{1i}^{*}\Phi_{2,1} + \cdots + \lambda_{ri}^{*}\Phi_{2,r}\right) \\ + \cdots + F_{rt}^{*}\left(\lambda_{1i}^{*}\Phi_{r,1} + \lambda_{2i}^{*}\Phi_{r,2} + \cdots + \lambda_{ri}^{*}\Phi_{r,r}\right) + e_{i,t}\end{aligned}$$

Given estimates of factors and loadings from the full information proxy, $\left(\tilde{F}^{*}, \tilde{\Lambda}^{*}\right)$, the proposed estimator of the rotated attention matrix proceeds in two steps:

1. For each $i \in 1, \ldots, N$ regress $T \times 1$ vector $\hat{X}_i$ on $T \times r$ matrix $\tilde{F}^{*}$.
2. For each $k \in 1, \ldots, r$ regress N coefficients corresponding to k^{th} factor from first step onto $N \times r$ matrix $\tilde{\Lambda}^{*}$. The resulting vector of coefficients are estimates of the k^{th} row of Φ.

B.0.0.1. *First stage coefficients:* Consider variable i. Replacing full information variables with estimated coefficients, the model implies:

$$\hat{X}_i = \tilde{F}^{*}\Phi\tilde{\lambda}_i^{*} + e_i$$

Let $\widehat{\left(\Phi\lambda_i^{*}\right)} = \left(\tilde{F}^{*\prime}\tilde{F}^{*}\right)^{-1}\tilde{F}^{*\prime}\hat{X}_i$. Regress both sides by the estimated factors:

$$\left(\tilde{F}^{*\prime}\tilde{F}^{*}\right)^{-1}\tilde{F}^{*\prime}\hat{X}_i = \Phi\tilde{\lambda}_i^{*} + \left(\tilde{F}^{*\prime}\tilde{F}^{*}\right)^{-1}\tilde{F}^{*\prime}e_i$$

$$\underbrace{\widehat{\left(\Phi\lambda_i^{*}\right)}}_{r\times 1} = \underbrace{\Phi}_{r\times r}\underbrace{\tilde{\lambda}_i^{*}}_{r\times 1} + \underbrace{\left(\tilde{F}^{*\prime}\tilde{F}^{*}\right)^{-1}\tilde{F}^{*\prime}e_i}_{r\times 1}$$

B.0.0.2. *Second stage coefficients* Take the k^{th}, $(k \in 1, \dots r)$ coefficient for each of the N forecasted variables (denoted by $\widehat{\left(\Phi\lambda_i^*\right)}_k$), and regress it on the full set of loadings.

$$\underbrace{\widehat{(\Phi\Lambda^{*\prime})'}_k}_{N\times 1} = \underbrace{\tilde{\Lambda}^*}_{N\times r}\underbrace{\Phi'_{k,1:r}}_{r\times 1} + \underbrace{e'}_{N\times T}\underbrace{\tilde{F_k}^*}_{T\times 1}\underbrace{\left(\tilde{F_k}^{*\prime}\tilde{F_k}^*\right)^{-1}}_{1\times 1}$$

$$\underbrace{\left(\tilde{\Lambda}^{*\prime}\tilde{\Lambda}^*\right)^{-1}\tilde{\Lambda}^{*\prime}\widehat{(\Phi\Lambda^{*\prime})'}_k}_{r\times 1} = \Phi'_{k,1:r} + \left(\tilde{\Lambda}^{*\prime}\tilde{\Lambda}^*\right)^{-1}\tilde{\Lambda}^{*\prime}e'\tilde{F_k}^*\left(\tilde{F_k}^{*\prime}\tilde{F_k}^*\right)^{-1}$$

Denote $\widehat{\widehat{(\Phi)}'}_{k,1:r} = \left(\tilde{\Lambda}^{*\prime}\tilde{\Lambda}^*\right)^{-1}\tilde{\Lambda}^{*\prime}\widehat{(\Phi\Lambda^{*\prime})'}_k$:

$$\widehat{\widehat{(\Phi)}'}_{k,1:r} - \Phi'_{k,1:r} = \left(\tilde{\Lambda}^{*\prime}\tilde{\Lambda}^*\right)^{-1}\tilde{\Lambda}^{*\prime}e'\tilde{F_k}^*\left(\tilde{F_k}^{*\prime}\tilde{F_k}^*\right)^{-1}$$

$$\sqrt{TN}\left(\widehat{\widehat{(\Phi)}'}_{k,1:r} - \Phi'_{k,1:r}\right) = \left(\frac{\tilde{\Lambda}^{*\prime}\tilde{\Lambda}^*}{N}\right)^{-1}\frac{\tilde{\Lambda}^{*\prime}e'\tilde{F_k}^*}{\sqrt{NT}}\left(\frac{\tilde{F_k}^{*\prime}\tilde{F_k}^*}{T}\right)^{-1}$$

By Assumption 8, the middle term converges to $\Sigma_{\Lambda eF_k}$. The term on the left converge to Σ_Λ^{-1}. Under the normalization of $\frac{F^{*\prime}F^*}{T} = \mathbb{I}$, the asymptotic distribution of the estimator for the k^{th} row of the rotated attention matrix Φ is:

$$\sqrt{TN}\left(\widehat{\widehat{(\Phi)}'}_{k,1:r} - \Phi'_{k,1:r}\right) \to N\left(0, \Sigma_\Lambda^{-1}\Sigma_{\Lambda eF_k}\Sigma_\Lambda^{-1}\right)$$

B.1. *Monte Carlo Simulations*

APPENDIX C: DATA ON MACROECONOMIC FORECASTS

The *Survey of Professional Forecasters* (SPF) is an anonymous quarterly survey of professional macroeconomic forecasters. I use quarterly data from 1984 until 2019 ($T = 144$) for forecasts made of the following set of variables: nominal GDP, real GDP, real Consumption, business fixed investment, residential investment, inventories, corporate profits, industrial production, employment (nonfarm payroll), unemployment, housing starts, GDP price index, CPI, real net exports, real government spending (federal and state and local), Treasury Bills (3 Month), Treasury Bonds (10 Year), Treasury Spreads (10 Year - 3 Month), and BAA corporate bond spreads over treasuries. Forecasts of the percent change, change (interest rates, exports, inventories, employment), or level (Treasury and BAA spreads) are

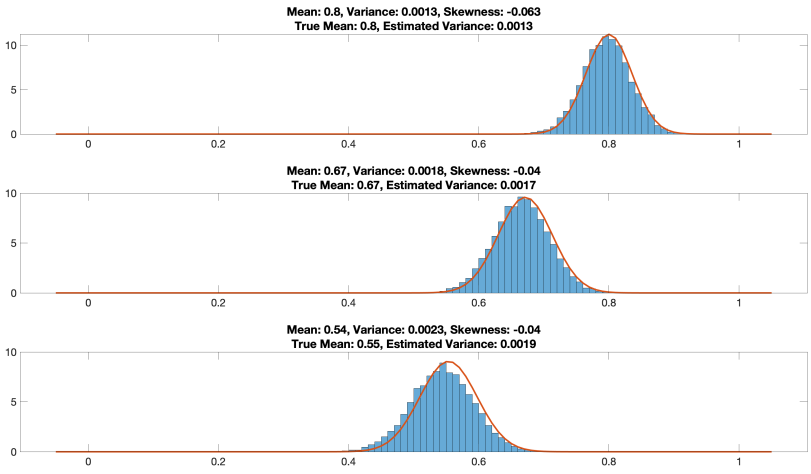


FIGURE B.1.—Monte Carlo Simulations with a Low Cost of Information

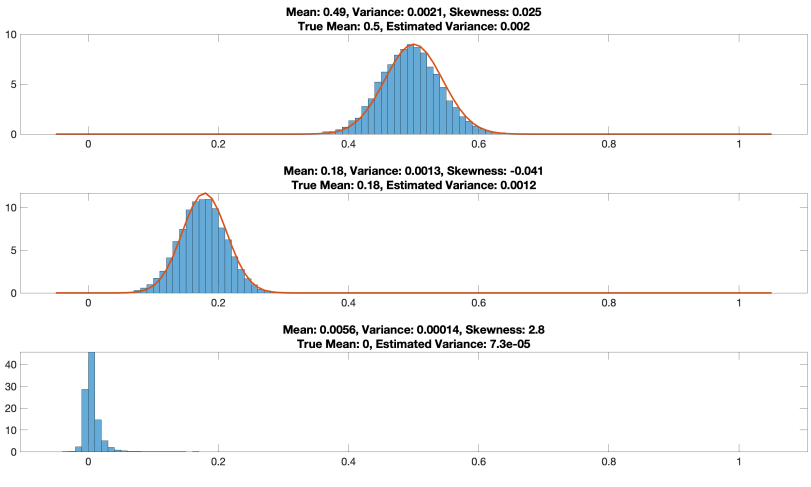


FIGURE B.2.—Monte Carlo Simulations with a High Cost of Information

used for each series. Forecasts are taken at 6 different horizons: current quarter $(t + 0)$,

subsequent four quarter ($t+h,\ h=1,2,3,4$), and annual forecasts for two calendar years ahead ($t+\overline{2Y}$).[14]

Individuals surveyed by the SPF have an identifier associated with them, but drop in and out of the sample and do not stay in the sample for the duration of the period under consideration. Following a similar method used by Giacoletti et al. (2021) and Bianchi et al. (2022), I build a time series of forecasts for forecasters corresponding to different percentile types based on forecast performance. A forecast at percentile j will include each of the 20 variables forecasted at 6 different horizons, constituting 120 forecasted variables for each time period. The forecast at time t made by percentile forecaster j will correspond to the entire vector of forecasts made by the individual forecaster who ranked at the percentile j in average forecast performance over the year prior against real-time observations of the data from the Real-Time Data Set for Macroeconomists from the Federal Reserve Bank of Phildelphia, with a higher percentile corresponding to a better (lower error) forecast. Average forecast performance is based on the average standardized forecast error for all variables made at the 5 quarterly horizons ($t+h,\ h=0,1,2,3,4$). The two-year ahead annual forecast is excluded. Forecaster rank is persistent.

In addition to forecasts made by forecasters of different performance rank, I consider a forecast comprised of the median forecasted value for each variable, which I will refer to as the Consensus forecast. The Consensus forecast is taken as the median forecasted value for each variable at each time horizon. Thus, the vector of Consensus forecasts at time t will not correspond to an individual forecaster, but to the median across all forecasters for each variable and horizon.

[14] Forecasts for percent changes and changes are for the annual average two calendar years ahead relative to the annual average one calendar year ahead. If a forecast is made in 2004, this will correspond to the forecast of 2006 relative to 2005, regardless of when in 2004 the forecast is made.

TABLE C.I

TOP 10 VARIABLE CORRELATION WITH FACTORS BY MARGINAL R^2

Consensus

Factor 1	R^2	Factor 2	R^2	Factor 3	R^2
RGDPB	0.782	EMP4	0.700	PGDP4	0.469
RGDP4	0.704	UNEMP3	0.654	PGDP5	0.461
NGDPB	0.677	EMP3	0.643	PGDP6	0.451
RGDP5	0.659	EMP5	0.580	PGDP3	0.384
INDPRODB	0.626	BONDB	0.525	CPI6	0.336
RCONSUMB	0.626	UNEMP2	0.446	CPI4	0.335
RGDP6	0.623	TSPREAD5	0.444	CPI5	0.334
RNRESIN6	0.613	EMPB	0.442	CPIB	0.316
INDPROD4	0.609	UNEMP4	0.431	CPI3	0.311
NGDP4	0.599	BOND4	0.425	PGDP2	0.294
Factor 4	R^2	Factor 5	R^2	Factor 6	
RRESINV6	0.420	BOND3	0.237	RSLGOV3	0.207
RRESINV5	0.383	BOND4	0.225	RSLGOV2	0.201
RRESINV4	0.306	BOND2	0.222	REXPORT5	0.193
HOUSING5	0.281	BOND5	0.222	REXPORT3	0.190
HOUSING3	0.281	BOND6	0.199	REXPORT4	0.185
HOUSING6	0.278	PGDP3	0.182	REXPORT6	0.173
HOUSING4	0.250	RGDP2	0.159	REXPORT2	0.148
HOUSINGB	0.240	HOUSING6	0.157	RSLGOV4	0.145
RRESINVB	0.231	INDPROD2	0.154	RFEDGOV6	0.141
RRESINV3	0.193	RCONSUM2	0.151	CPI3	0.137

Source: *Survey of Professional Forecasters*, Author calculations. The tables show the top 10 variables (ordered) by marginal correlation with the first six principal components of the Consensus Forecast. The mneumonics are the following: RGDP: Real GDP; RCONSUM: Real Consumption; NGDP: Nom. GDP; RRESINV: Residential Investment; RNRESIN: NonRes Inv.; INDPROD: Industrial Production; HOUSING: Housing Starts; BOND: AAA - T3M; TSPREAD: T10Y-T3M; BOND: AAA - T3M3; EMP: Employment; PGDP: GDP Price Index; CPI: CPI; CPROF: Corp. Profits; RFEDGOV: Real Fed Gov; RCBI: Real Inventories; UNEMP: Unemployment; TBOND: T10Y; TBILL: T3M. The number at the end corresponds to the horizon forecast, 2 is the current quarter, 3 the subsequent quarter, and so on. "B" denotes the 2-year ahead annual average forecast.

TABLE C.II

TOP 10 VARIABLE CORRELATION WITH FACTORS BY MARGINAL R^2

90th Percentile

Factor 1	R^2	Factor 2	R^2	Factor 3	R^2
RGDPB	0.763	BOND3	0.545	CPI6	0.305
RCONSUMB	0.669	BOND4	0.524	CPI5	0.296
RGDP4	0.633	BOND2	0.507	TSPREAD6	0.296
NGDPB	0.607	BOND5	0.489	CPIB	0.284
NGDP5	0.603	TSPREAD3	0.431	CPI4	0.277
NGDP4	0.596	EMP3	0.404	CPI3	0.274
EMP6	0.590	TSPREAD4	0.395	TSPREAD5	0.273
RCONSUM4	0.577	BOND6	0.375	PGDP5	0.273
RGDP5	0.577	UNEMP2	0.372	PGDP4	0.260
RCONSUM5	0.562	EMP2	0.352	PGDP6	0.252
Factor 4	R^2	Factor 5	R^2	Factor 6	
CPI5	0.293	CPROF6	0.302	TBOND2	0.297
CPI6	0.278	RCONSUM2	0.262	RRESINV4	0.230
CPI4	0.278	TSPREAD6	0.228	TBILL3	0.228
RFEDGOV6	0.257	INDPROD6	0.228	RRESINV3	0.213
CPROFB	0.233	RNRESIN2	0.219	RRESINV5	0.208
HOUSINGB	0.226	RNRESIN3	0.190	EMP4	0.162
PGDP6	0.224	TBOND2	0.187	RRESINVB	0.159
CPROF3	0.220	TSPREAD5	0.184	EMP2	0.154
PGDP5	0.212	TBONDB	0.170	HOUSING2	0.141
CPI3	0.210	PGDP3	0.161	TSPREAD2	0.134

Source: *Survey of Professional Forecasters*, Author calculations. The tables show the top 10 variables (ordered) by marginal correlation with the first six principal components of the 90th percentile forecaster. The mneumonics are the following: RGDP: Real GDP; RCONSUM: Real Consumption; NGDP: Nom. GDP; RRESINV: Residential Investment; RNRESIN: NonRes Inv.; INDPROD: Industrial Production; HOUSING: Housing Starts; BOND: AAA - T3M; TSPREAD: T10Y-T3M; BOND: AAA - T3M3; EMP: Employment; PGDP: GDP Price Index; CPI: CPI; CPROF: Corp. Profits; RFEDGOV: Real Fed Gov; RCBI: Real Inventories; UNEMP: Unemployment; TBOND: T10Y; TBILL: T3M. The number at the end corresponds to the horizon forecast, 2 is the current quarter, 3 the subsequent quarter, and so on. "B" denotes the 2-year ahead annual average forecast.

TABLE C.III

TOP 10 VARIABLE CORRELATION WITH FACTORS BY MARGINAL R^2

10th Percentile

Factor 1	R^2	Factor 2	R^2	Factor 3	R^2
RGDPB	0.733	TSPREAD5	0.677	TBOND4	0.467
RCONSUMB	0.566	TSPREAD4	0.657	PGDP4	0.355
RGDP5	0.528	TSPREAD3	0.627	RCBI6	0.302
NGDPB	0.508	TSPREAD2	0.541	RCONSUM4	0.292
NGDP5	0.475	TSPREAD6	0.470	REXPORT6	0.277
RRESINVB	0.456	BOND6	0.455	PGDP6	0.272
RRESINV5	0.452	BOND4	0.447	RNRESIN6	0.234
INDPRODB	0.431	BOND5	0.444	INDPROD6	0.233
NGDP6	0.403	BOND3	0.394	CPIB	0.202
RNRESINB	0.402	BONDB	0.387	RGDP3	0.198
Factor 4	R^2	Factor 5	R^2	Factor 6	
RFEDGOV3	0.479	CPIB	0.288	PGDP6	0.323
RFEDGOVB	0.478	CPI4	0.281	EMP6	0.257
RFEDGOV6	0.477	CPI5	0.276	UNEMP3	0.247
RFEDGOV4	0.455	RFEDGOV5	0.248	REXPORT3	0.237
RFEDGOV5	0.376	RFEDGOV4	0.246	UNEMP2	0.192
RFEDGOV2	0.373	RFEDGOV2	0.231	RNRESIN3	0.190
INDPROD4	0.353	TBOND5	0.198	PGDP4	0.189
INDPROD5	0.225	TBONDB	0.193	EMP5	0.189
RSLGOV4	0.190	RFEDGOV6	0.188	EMP3	0.154
NGDP2	0.180	RCONSUM3	0.180	RGDP2	0.143

Source: *Survey of Professional Forecasters*, Author calculations. The tables show the top 10 variables (ordered) by marginal correlation with the first six principal components of the 10th percentile forecaster. The mneumonics are the following: RGDP: Real GDP; RCONSUM: Real Consumption; NGDP: Nom. GDP; RRESINV: Residential Investment; RNRESIN: NonRes Inv.; INDPROD: Industrial Production; HOUSING: Housing Starts; BOND: AAA - T3M; TSPREAD: T10Y-T3M; BOND: AAA - T3M3; EMP: Employment; PGDP: GDP Price Index; CPI: CPI; CPROF: Corp. Profits; RFEDGOV: Real Fed Gov; RCBI: Real Inventories; UNEMP: Unemployment; TBOND: T10Y; TBILL: T3M. The number at the end corresponds to the horizon forecast, 2 is the current quarter, 3 the subsequent quarter, and so on. "B" denotes the 2-year ahead annual average forecast.

APPENDIX D: COST FUNCTION PARAMETER ESTIMATION DETAILS

I estimate the parameters ω_j and α_j for each type of forecaster using minimum distance. The eigenvalues from the full information benchmark forecast, $\lambda_1, \lambda_2, \ldots \lambda_k$, are taken as known. Let $\widehat{1-\Omega_j^i}$ represent the estimated value of the i^{th} diagonal value of the attention matrix. Each diagonal element provides hypothesis: $\left(\widehat{1-\Omega_j^i}\right) - \left(1-\left(\frac{\omega_j}{\lambda_i}\right)^{\alpha_j}\right)$. In the estimation procedure I use a log transformation.

In order to jointly estimate the two parameters, I use the expectations maximizing algorithm. In particular I start with initial values for α_j, ω_j, and the weight matrix W denoted by α_j^0, ω_j^0, and W^0, which are 1, 1 and the identity matrix, respectively. I then iterate between the following two steps until the estimated values of α_j and ω_j converge.[15]

1. Given α_j^0 and W^0 estimate $\hat{\omega}_j$ minimize the following vector norm:

$$\hat{\omega}_j = \underset{\omega_j}{\text{argmin}} \begin{pmatrix} \log\left(1-\left(\widehat{1-\Omega_j^1}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_1}\right) \\ \log\left(1-\left(\widehat{1-\Omega_j^2}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_2}\right) \\ \cdots \\ \log\left(1-\left(\widehat{1-\Omega_j^k}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_k}\right) \end{pmatrix}' W^0 \begin{pmatrix} \log\left(1-\left(\widehat{1-\Omega_j^1}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_1}\right) \\ \log\left(1-\left(\widehat{1-\Omega_j^2}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_2}\right) \\ \cdots \\ \log\left(1-\left(\widehat{1-\Omega_j^k}\right)\right) - \alpha_j^0 \log\left(\frac{\omega_j}{\lambda_k}\right) \end{pmatrix}$$

2. Let $\omega_j^0 = \hat{\omega}_j$ from step 1. Use this value and W^0 to estimate $\hat{\alpha}_j$ by minimizing the following:

$$\hat{\alpha}_j = \underset{\alpha_j}{\text{argmin}} \begin{pmatrix} \log\left(1-\left(\widehat{1-\Omega_j^1}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_1}\right) \\ \log\left(1-\left(\widehat{1-\Omega_j^2}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_2}\right) \\ \cdots \\ \log\left(1-\left(\widehat{1-\Omega_j^k}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_k}\right) \end{pmatrix}' W^0 \begin{pmatrix} \log\left(1-\left(\widehat{1-\Omega_j^1}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_1}\right) \\ \log\left(1-\left(\widehat{1-\Omega_j^2}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_2}\right) \\ \cdots \\ \log\left(1-\left(\widehat{1-\Omega_j^k}\right)\right) - \alpha_j \log\left(\frac{\omega_j^0}{\lambda_k}\right) \end{pmatrix}$$

Replace α_j^0 with $\hat{\alpha}_j$. Replace W^0 with the inverse of the empirical variance of the moment conditions.[16]

[15] Estimated parameters converge after 2 or 3 iterations.

[16] The variance matrix is scaled by the mean to avoid inverting a matrix with eigenvalues close to 0.

The resulting values capture the implied marginal cost and curvature of the information cost function for forecasters of different rank.

Given convergence of the estimated values of (α_j, ω_j), the asymptotic variance of the estimator, $\mathcal{V}$ is given by the following:

$$\mathcal{V} = \left(DWD'\right)^{-1}$$

Where D is the gradient of the moments with respect to the elements of (α_j, ω_j) evaluated at $(\hat{\alpha}_j, \hat{\omega}_j)$. Here the matrix D is given by:

$$D = \begin{pmatrix} \log\left(\frac{\omega_j}{\lambda_1^B}\right)\frac{\alpha_j}{\omega_j} \\ \log\left(\frac{\omega_j}{\lambda_2^B}\right)\frac{\alpha_j}{\omega_j} \\ \vdots \quad\quad \vdots \\ \log\left(\frac{\omega_j}{\lambda_k^B}\right)\frac{\alpha_j}{\omega_j} \end{pmatrix}$$

Standard errors in the empirical results are computed using this estimated variance matrix.

TABLE D.I

ESTIMATES OF ω_j AND α_j BY FORECAST PERFORMANCE

	10th	20th	30th	40th	50th	60th	70th	80th	90th
ω_j	1.124	0.972	0.847	0.983	0.764	0.801	0.717	0.792	0.774
SE	(0.020)	(0.006)	(0.017)	(0.026)	(0.027)	(0.021)	(0.027)	(0.011)	(0.002)
α_j	0.511	0.696	0.757	0.718	0.815	0.816	0.839	0.837	0.854
SE	(0.010)	(0.010)	(0.017)	(0.023)	(0.029)	(0.017)	(0.027)	(0.010)	(0.002)

Source: Federal Reserve Bank of Philadelphia Real-Time Data Set for Macroeconomists, *Survey of Professional Forecasters*, Federal Reserve Board Selected Interest Rates (H.15), Federal Reserve Bank of St. Louis, Moody's, Author calculations. Estimates of α_j and ω_j by forecaster rank with standard errors.

REFERENCES

AFROUZI, HASSAN AND CHOONGRYUL YANG (2021): "Dynamic Rational Inattention and the Phillips Curve," *Working Paper.* [4, 5, 6, 12, 16]

ANDRADE, PHILIPPE, RICHARD K. CRUMP, STEFANO EUSEPI, AND EMANUEL MOENCH (2016): "Fundamental disagreement," *Journal of Monetary Economics*, 83, 106–128. [5]

BAI, JUSHAN AND SERENA NG (2002): "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70 (1), 191–221. [15]

——— (2013): "Principal components estimation and identification of static factors," *Journal of Econometrics*, 176 (1), 18–29. [27]

——— (2019): "Rank regularized estimation of approximate factor models," *Journal of Econometrics*, 212 (1), 78–96, big Data in Dynamic Predictive Econometric Modeling. [15, 16]

BARTOŠ, VOJTĚCH, MICHAL BAUER, JULIE CHYTILOVÁ, AND FILIP MATĚJKA (2016): "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 106 (6), 1437–75. [4]

BIANCHI, FRANCESCO, SYDNEY C. LUDVIGSON, AND SAI MA (2022): "Belief Distortions and Macroeconomic Fluctuations," *American Economic Review*, 112 (7), 2269–2315. [22, 38]

CAI, JIAN-FENG, J. CANDÈS, EMMANUEL, AND ZUOWEI SHEN (2010): "A Singular Value Thresholding Algorithm for Matrix Completion," *SIAM Journal on Optimization*, 20 (4), 1956–1982. [16]

CAPLIN, ANDREW, MARK DEAN, AND JOHN LEAHY (2019): "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *The Review of Economic Studies*, 86 (3 (308)), pp. 1061–1094. [4]

COIBION, OLIVIER AND YURIY GORODNICHENKO (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105 (8), 2644–78. [5]

COVER, THOMAS AND JOY THOMAS (2006): *Elements of Information Theory*, John Wiley and Sons, second ed. [3]

DEAN, MARK AND NATHANIEL NATHANIEL NELIGH (Forthcoming): "Experimental Tests of Rational Inattention," *Journal of Political Economy*. [4]

DOVERN, JONAS (2015): "A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data," *European Economic Review*, 80, 16–35. [5]

DURBIN, J. AND S. J. KOOPMAN (2012): *Time Series Analysis by State Space Methods*, Oxford University Press, second ed. [33]

FULTON, CHAD (2017): "Mechanics of Linear Quadratic Gaussian Rational Inattention Tracking Problems," *Finance and Economics Discussion Series*. [4, 5]

GIACOLETTI, MARCO, KRISTOFFER T. LAURSEN, AND KENNETH J. SINGLETON (2021): "Learning From Disagreement in the U.S. Treasury Bond Market," *The Journal of Finance*, 76 (1), 395–441. [21, 38]

GIANNONE, DOMENICO, LUCREZIA REICHLIN, AND DAVID SMALL (2008): "Nowcasting: The real-time informational content of macroeconomic data," *Journal of Monetary Economics*, 55 (4), 665–676. [6, 17]

HÉBERT, BENJAMIN AND JENNIFER LA'O (2023): "Information Acquisition, Efficiency, and Non-Fundamental Volatility," *Journal of Political Economy*, 131 (10), 2666–2723. [2]

HÉBERT, BENJAMIN AND MICHAEL WOODFORD (2021): "Neighborhood-Based Information Costs," *American Economic Review*, 111 (10), 3225–55. [2, 4, 13]

HERBST, EDWARD P. AND FABIAN WINKLER (2021): "The Factor Structure of Disagreement," *Board of Governors of the Federal Reserve System (U.S.)*. [5, 6]

KHAW, MEL WIN, ZIANG LI, AND MICHAEL WOODFORD (2020): "Cognitive Imprecision and Small-Stakes Risk Aversion," *The Review of Economic Studies*, 88 (4), 1979–2013. [4]

KŐSZEGI, BOTOND AND FILIP MATĚJKA (2020): "Choice Simplification: A Theory of Mental Budgeting and Naive Diversification*," *The Quarterly Journal of Economics*, 135 (2), 1153–1207. [4, 5, 6]

MACKOWIAK, BARTOSZ AND MIRKO WIEDERHOLT (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99 (3), 769–803. [4]

MANKIW, N. GREGORY, RICARDO REIS, AND JUSTIN WOLFERS (2003): "Disagreement about Inflation Expectations," *NBER Macroeconomics Annual*, 18, 209–248. [5]

MATĚJKA, FILIP (2015): "Rationally Inattentive Seller: Sales and Discrete Pricing," *The Review of Economic Studies*, 83 (3), 1125–1155. [4]

MATĚJKA, FILIP AND ALISDAIR MCKAY (2015): "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review*, 105 (1), 272–98. [3, 4]

MONDRIA, JORDI (2010): "Portfolio Choice, Attention Allocation, and Price Comovement," *Journal of Economic Theory*, 145, 1837–1864. [4]

SIMS, CHRISTOPHER A. (1998): "Stickiness," *Carnegie-Rochester Conference Series on Public Policy*, 49, 317–356. [3]

——— (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50 (3), 665–690, swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information. [2, 3, 5]

STEINER, JAKUB, COLIN STEWART, AND FILIP MATĚJKA (2017): "Rational Inattention Dynamics: Inertia and Delay in Decision-Making," *Econometrica*, 85 (2), 521–553. [4]

STOCK, JAMES H. AND MARK W. WATSON (2002a): "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97 (460), 1167–1179. [15]

——— (2002b): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20 (2), 147–162. [15]

WOODFORD, MICHAEL (2009): "Information-constrained state-dependent pricing," *Journal of Monetary Economics*, 56, S100–S124, supplement issue: December 12-13, 2008 Research Conference on 'Monetary Policy under Imperfect Information' Sponsored by the Swiss National Bank (http://www.snb.ch) and Study Center Gerzensee (www.szgerzensee.ch). [4]

YANG, MING (2015): "Coordination with flexible information acquisition," *Journal of Economic Theory*, 158, 721–738, symposium on Information, Coordination, and Market Frictions. [4]