Attention Allocation and the Factor Structure of Forecasts

Susannah Scanlan*
Columbia University

Click Here For Current Version

December 30, 2023

Abstract

I explore how forecaster attention, or the degree to which new information is incorporated into forecasts, is reflected at the lower-dimensional factor representation of multivariate forecast data. When information is costly to acquire, forecasters may pay more attention to some sources of information and ignore others. How much attention they pay will determine the strength of the forecast correlation (factor) structure. Using a factor model representation, I show that a forecast made by a rationally inattentive agent will include an extra shrinkage and thresholding "attention matrix" relative to a full information benchmark, and propose an econometric procedure to estimate it. I show that the mapping from theoretical attention allocation to factor model representation is valid for a broad class of information cost functions. Differences in the degree of forecaster attentiveness can explain observed differences in empirical shrinkage in professional macroeconomic forecasts relative to a consensus benchmark. Better-performing forecasters have higher measured attention (lower shrinkage) than their poorly-performing peers. Measured forecaster attention to multiple dimensions of the information space can largely be captured by a single scalar cost parameter.

*Columbia University, Department of Economics. Email: ss5605@columbia.edu. I would like to thank Michael Woodford, Serena Ng, Jennifer L'aO, Hassan Afrouzi, and the faculty at Columbia University for their helpful guidance. I would additionally like to thank Michelle Jiang, Shin Oblander, Jeffrey Guo, Qianyan Zhang and Wendy Morrison for useful comments and suggestions.

1 Introduction

Forecasts are important. Expectations for the future drive economic decisions today, and so understanding how these expectations are formed is key for optimal policymaking. Professional macroeconomic forecasts exhibit large dispersion across both time and variable, suggesting that professionals have different processes for making their forecasts.

A very specific way in which professional macroeconomic forecasts differ occurs at a lower dimensional representation. Empirically, panels of macroeconomic forecasts from the *Survey of Professional Forecasters* corresponding to forecasters of different performance rank have similar reduced form models, but differ in the degree to which relevant information gets mapped into forecasts[1]. A lower mapping of relevant information into a forecast results in shrinkage at the lower dimension relative to a consensus benchmark forecast, or a forecast comprised of the median forecasted value of each variable. Shrinkage refers to the average ratio of coefficients on common factors of forecasted variables from the professional forecasts and the consensus forecast.

While specific, when analyzed through the lens of a rational inattention model this observed shrinkage can be interpreted as differences in the degree to which forecasters are paying attention to available data. I derive a novel mapping between the theoretical information choice and statistical factor model frameworks. Under this mapping, shrinkage in the forecast factor structure results from lower attention by the forecaster.

There is a weaker mapping (shrinkage) from latent common factors into the professional forecast panel data than in the consensus forecast - common factors corresponding to innovations in real growth, inflation and interest rate spreads are less important in professional forecast data than in the consensus benchmark. The degree of shrinkage is both increasing as forecaster performance declines, and increasing along factors that explain less variance (less important). Additionally, statistical tests indicate fewer common factors in forecasts from professionals relative to the consensus forecast, which I will refer to as "thresholding".

The information that someone pays attention to is an important determinant in their forecasts. Consider someone making a prediction of next month's inflation in the United States. They could choose a number at random, but this probably would not be a very good forecast. They could instead base their forecast on their knowledge of last month's CPI report, or their own recent shopping experiences. The accumulation of inflation signals that the forecaster acquires or pays attention to will determine their set of information. If the person acquires more signals, they will have a more accurate idea of the levels or trends in price changes and will likely make a better

[1]Data include 20 variables made over 6 forecast horizons are considered over the period from 1984 to 2019. The reduced form models correspond to the factor structure estimated via principal components. Under a factor model, variables in a panel of data can be represented as weighted combinations of latent common factors, which capture maximal covariance in descending order, plus idiosyncratic error.

forecast compared to someone who selects a number at random.

A person's degree of attentiveness, or the number and quality of signals they collect, will depend on how costly it is for them to gather or process information. If someone faced a high cost on or a low capacity for processing information[2], it could be optimal for them to not pay any attention and select a number at random for their inflation forecast. Someone facing a low enough cost would pay some attention, but will face a decision of the optimal time to stop gathering information. They could read just the headline numbers of recent inflation reports, or they could go through the details of the report. They could even purchase data from a private vendor selling proprietary data on high-frequency price changes. There is an endless amount of signals the person could acquire. At some point, however, they would likely determine the cost to not be worth the benefit.

The optimal information choice of someone forecasting many variables will reflect the increased dimensionality. The forecaster would pay more attention to information relevant to many of the variables than to information that might only affect a few. Suppose in addition to inflation, the forecaster was asked to forecast the growth in real GDP, nominal imports, and housing starts in the next month and over the next year. What information would the person pay attention to? All four of the variables are related to aggregate economic conditions, so they would likely pay attention to that - if the economy is doing well GDP growth will be high, people will purchase more imported goods, will be more likely to purchase houses and in turn spur inflation. What about acquiring information on exchange rates? Exchange rates could have an effect on imports and inflation through changes in prices, but are not likely to have a big impact on housing starts or GDP growth. A forecaster would probably decide to pay less attention, and possibly no attention, to exchange rate data than to data on aggregate growth. In the multivariate forecast setting, the benefit of information is aggregated across multiple variables.

The attention allocation of forecasters in a multivariate setting induces a correlation structure between forecasted variables. If a forecaster decides to pay attention to exchange rates, their forecasts for imports and inflation will display a greater correlation through this connection to exchange rates. All of the forecasted variables will be correlated through the aggregate economic conditions channel. If the forecaster looks at interest rates, longer-term forecasts for all of the variables will show greater correlation through the interest rate channel. The degree of attention by forecasters to different sources of information (e.g. aggregate activity, exchange rate, interest rate) will determine the strength of correlation in forecasted variables along that channel.

In forecasts made repeatedly over time this correlation structure will be reflected in the latent

[2]The term "cost" is general and captures both acquisition and processing costs. It could capture a cognitive processing capacity, an actual cost of purchasing a set of information, or the shadow cost of marginal attention in the case that a person faces a constraint on information. Under a capacity constraint, the "cost" of information reflects the shadow cost of additional information acquisition in the Lagrangian.

factor structure of the panel data[3]. Factor models provide a method of representing a large number of variables in terms of common factors that capture most of the variation in the dataset. A single variable in the panel dataset can be represented as a weighted combination of unobserved common factors plus a variable-specific idiosyncratic component. The weights are represented by a loading matrix. If the factors are estimated via principal components, they represent the largest sources of co-movement in the data in descending order of importance[4].

Less attention from a forecaster will lead to a weaker correlation structure in forecast data, which can be represented by a shrinkage and thresholding matrix applied to the loadings. Someone paying less attention than someone else will have a greater degree of shrinkage in the loading matrix (lower magnitude weights). If they decide to ignore information (thresholding) there will be fewer common factors in the lower-dimensional representation of their forecast. Relative to a full attention, or information, benchmark, any forecaster with a constraint on information processing should display shrinkage.

I formalize the mapping from an agent's optimal choice of attention to information to the implied factor structure of their forecasts. To model the forecaster's information choice, I take as a starting point a rational inattention model in a linear quadratic Gaussian setting. A forecaster acquires costly signals about a set of latent shocks, potentially as large as the set of variables being forecasted, and makes a forecast based on their belief about the realization of the shocks. I derive the forecaster's belief about the states, and hence forecasts, for a new and generalized set of cost functions, which I call Separable Spectral. I then take the forecasts made by a theoretical rationally inattentive agent repeatedly over time as the data generating process for a statistical factor model. The rationally inattentive forecast can be written as the factors and loadings estimated from the full information forecast with an additional shrinkage and thresholding matrix parameterized by a single information cost term. I describe the elements of this shrinkage and thresholding matrix as "attention" to different dimensions of the space of information.

Returning to professional macroeconomic forecasts, the observed shrinkage in forecast data made by professionals relative to a consensus benchmark can be interpreted as the relative attention of forecasters. Better performing forecasters pay more attention than poorly performing forecasters, and so display less shrinkage. All types of forecasters pay relatively less attention to less important factors, and so shrinkage is increasing in the factor order. The theoretical model additionally implies that that this multivariate attention can be captured by a single cost parameter, which I estimate. Top performing forecasters have a significantly lower measured cost than poorly performing forecasters.

[3] N variables forecasted T times.

[4] In the earlier example, aggregate economic activity would capture the largest amount of variance in the forecast dataset.

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.

Related papers that more specifically look at the factor structure of forecast data include Herbst and Winkler (2021) and Dovern (2015). These papers consider the factor structure of disagreement in forecasts, and assume exogenously-bestowed heterogeneity in signal observation error. In my paper, differences in modeled observation error arise from microfounded endogenous information choice. Andrade et al. (2016) additionally look at a multivariate forecast, but also assume exogenous differences in signal error.

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 Nowcaster, as in the dynamic factor model setup of Giannone et al. (2008). 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. Herbst and Winkler (2021) noted the connection between noisy information models and dynamic factor models, but did not include the endogenous choice of signal observation error. The endogenous information choice leads to predictions about the intensity of attention along more and less important dimensions. Previous papers have also pointed out the shrinkage and thresholding (Afrouzi and Yang (2021); Kőszegi and Matějka (2020)), albeit under terms not typically associated with the factor structure of panel data.

By explicitly connecting the rational inattention decision to a factor structure, and to principal component analysis more specifically, I can make use of econometric techniques from factor analysis[5]. This provides a new framework to examine the multidimensional predictions from a rational inattention model - in particular that attention is lower for less important factors, and that this decrease can be parameterized by a single information cost.

This paper builds on the literature on cost functions in linear quadratic Gaussian rational inattention models by introducing a new generalized cost funtion - separable spectral. The form of a cost function will determine how attention changes with dimension importance. Separable spectral cost functions are defined along the the eigenvalues of the variance matrix, which is fundamentally

[5]Bai and Ng (2002, 2006, 2013, 2019); Stock and Watson (2002a,b); Bai (2003); Cahan et al. (2023).

related to the mathematics of dimension reduction[6]. Rationally inattentive forecast models with different types of cost functions that fall under the umbrella of separable spectral can be mapped to shrinkage in the panel data of forecasts in the same way, allowing for a comparison of different model forms.

A common goal of policymakers is to set expectations of forecasters, indicating minimal dispersion. When viewed through the lens of a multivariate rational inattention model, it might not be feasible to achieve this. Policymakers cannot realistically change the cost or shadow cost of acquiring and processing more information. Additionally, measures of relative attention to different dimensions of the information space can be viewed as sources of risk, as someone's attention can shift quickly given the environment. Long-term expectations could appear anchored because forecasters optimally decide not to pay attention to things that affect forecasts of these variables, but could shift swiftly if circumstances changed.

Section 2 begins by presenting evidence that forecasts share a similar reduced form model across different types of forecasters, before presenting evidence of shrinkage of professional forecasts along the factor dimensions of the consensus forecast. Section 3 presents a rational inattention model that can explain this shrinkage with the generalized separable spectral cost functions. Section 4 takes the implied forecasts from the rational inattention model and maps them into a factor model. An estimator of the shrinkage, or relative attention, is presented. Section 5 returns to the empirical evidence from the *Survey of Professional Forecasters*.

2 Empirical Evidence of Shrinkage

In this section I present empirical evidence that macroeconomic forecasters of different types share the same reduced-form forecast model. I then define shrinkage relative to a benchmark forecast and show evidence of shrinkage in forecasts made by individuals relative to Consensus estimates.

2.1 Data on Macroeconomic Forecasts

The *Survey of Professional Forecasters* (SPF) is an anonymous quarterly survey of professional macroeconomic forecasters. Forecasters are asked to give predictions for a number of variables at various horizons.

I consider forecasted variables made from 1984 until 2019 for the following set of variables: nominal GDP, real GDP, real Consumption, business fixed investment, residential investment,

[6]The spectral theorem states that a single symmetric matrix can be decomposed into three matrices: its eigenvectors, a diagonal matrix of eigenvalues, and the transpose of the eigenvectors. Estimation of principal components relies on the min-max theorem, or Courant-Fisher-Weyl min-max principal, which characterizes the solution for a set of quadratic optimization problems (which minimizing forecast error falls into) in terms of the spectral decomposition.

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 (3Mo), Treasury Bonds (10Y), Treasury Spreads, 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 ($t+\overline{2Y}$). In the last case, 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.

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.

$$
\begin{aligned}
\hat{X}_t^j &= \left[\hat{x}_{1,t}^j, \hat{x}_{1,t+1}^j, \hat{x}_{1,t+2}^j \ldots, \hat{x}_{1,t+\overline{2Y}}^j, \hat{x}_{2,t}^j, \hat{x}_{2,t+1}^j, \ldots, \hat{x}_{20,t}^j, \hat{x}_{20,t+1}^j, \ldots \hat{x}_{20,t+\overline{2Y}}^j\right]' \\
&= \left[\hat{x}_{1,t}^j, \hat{x}_{2,t}^j, \hat{x}_{3,t}^j, \ldots \hat{x}_{120,t}^j\right]'
\end{aligned}
$$

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, 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.

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. The Consensus forecast performs on average 8% better than the 90th percentile type, and 18% better than the 10th percentile type. Over the timeframe considered, it has a better performance for 90% of the variables compared to the 90th percentile type, and 97.5% compared to the 10th percentile types.

2.2 Similarity of Estimated Factors Across Forecaster Types

The factors spanning panels from different forecaster types are correlated with the same sets of variables, and are shown to span the same space. Panels of forecasts corresponding to different performance rank display a similar lower-dimensional factor representation.

Factor Model Estimation

I consider panels of forecasts made by percentile ranks $j = 10, 90$ (these correspond to 10th and 90th percentile types by performance, 90th percentile better performing than 10th), as well as the Consensus forecast, which I call type $j = C$. For each panel of forecasts I first orthogonalize forecasted variables to the real-time realized value of the data. For each of the 20 variables (indexed by v) take the residuals:

$$\hat{x}^j_{v,t+h} = \rho_h x_{v,t-1} + e^j_{v,h}$$

The resulting panel of forecasts is composed of $e^j_{v,h}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Before estimating factors via principal components, the residuals $e^j_{v,h}$ are demeaned and standardized. Denote the resulting residualized, demeaned and standardized forecasts as

$$\tilde{X}^j_t = \left[e^j_{1,0}, e^j_{1,1}, \ldots e^j_{20,\overline{2Y}}\right]'$$

Factors $\tilde{F}$ and loadings $\tilde{\Lambda}$ from the panel of forecasts corresponding to type j are estimated via principal components as in Bai and Ng (2002):

$$\tilde{X}^j_t = \tilde{\Lambda}^j \tilde{F}^j_t + u^j_t$$

As a first exercise, I look at the correlation between estimated factors and forecasted variables *within forecaster*. For each forecaster type $j = 10, 90, C$, I regress each residualized forecasted variable $e^j_{v,h}$ on each factor and compute the R^2 from the regression.

Figures 1, 2, and 3 display the marginal R^2 by variable on the first three factors for the SPF Consensus forecast (in black) and types 90 (light blue, top panel) and 10 (gray, bottom panel). As can be seen from the figure, the bars largely line up along the same variables for the forecaster type. The first latent factor is correlated broadly with forecasts of economic activity. The second factor is correlated with Treasury and corporate bond spreads, as well as 10 Year and 3 Month Treasuries, and the third factor with inflation.

Figure 1: Marginal R^2 of Forecasted Variables on Factor 1

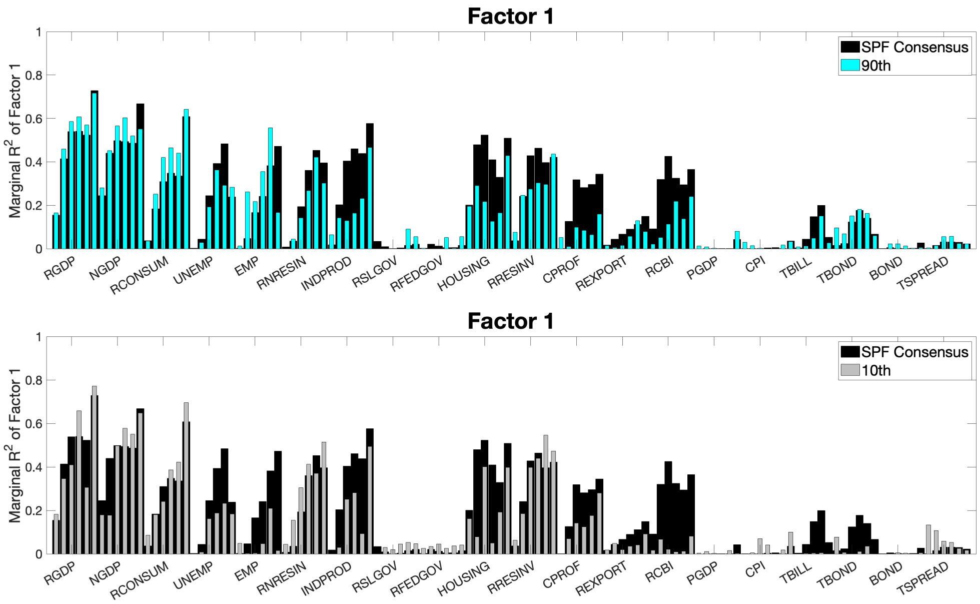


Figure shows the R^2 from the regression: $e^j_{v,h,t} = \hat{\beta}^j_{v,h,1}\tilde{F}^j_{1,t} + u^j_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$.
Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

Figure 2: Marginal R^2 of Forecasted Variables on Factor 2

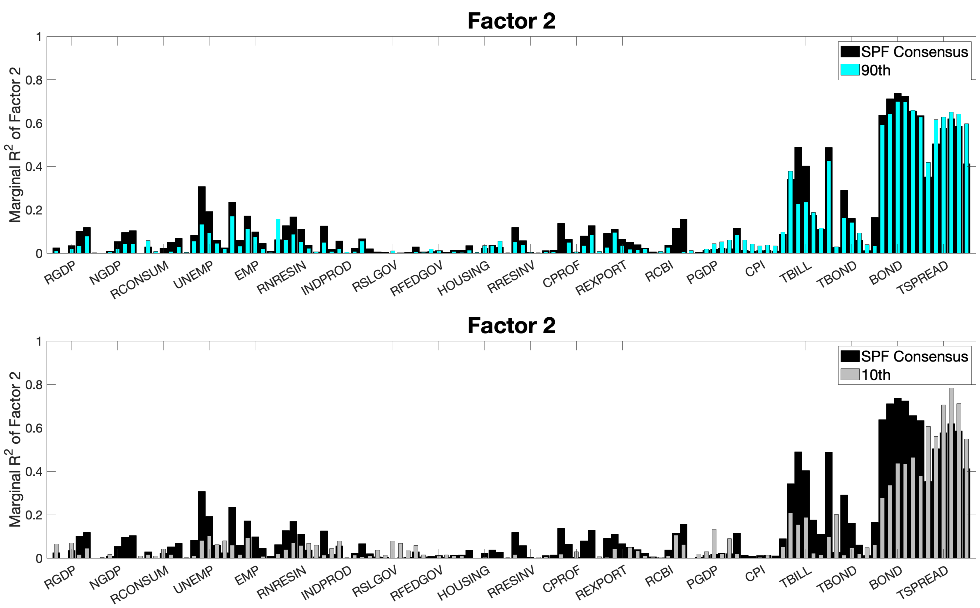


Figure shows the R^2 from the regression: $e^j_{v,h,t} = \hat{\beta}^j_{v,h,2}\tilde{F}^j_{2,t} + u^j_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$.
Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

Correlation with the same set of variables does not indicate that the estimated factors from each forecast panel are the same. To assess whether the factors span the same space, I consider canonical correlations between the estimated factors of types 10, 30, 50, 70 and 90 to the estimated

Figure 3: Marginal R^2 of Forecasted Variables on Factor 3

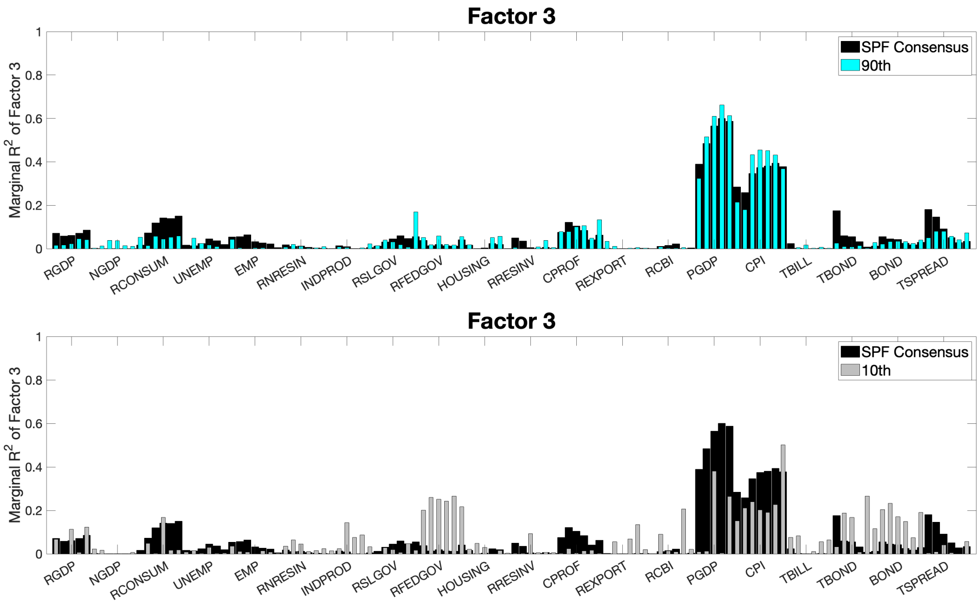


Figure shows the R^2 from the regression: $e^j_{v,h,t} = \hat{\beta}^j_{v,h,3}\tilde{F}^j_{3,t} + u^j_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon. The 4th factor is shown for the Consensus forecast. The 3rd and 4th factors explained a very similar share of variance in the Consensus forecast (7.88% versus 7.29%), but the 4th is shown as it corresponds to the same groupings of variables as the 10th and 90th rank forecasts.

factors from the Consensus. As can be seen in Table 1, there are significant canonical relationships between the pairs of estimated factors. The estimated factors largely span the same space.

2.3 Similarity of Estimated Loadings Across Forecaster Types

As a second excercise, I compare the loadings of forecasts made by different percentile forecasters on the estimated factors from the Consensus forecast. I run the following regressions:

$$e^j_{v,h,t} = \hat{\beta}^{j,C}_{v,h,f}\tilde{F}^C_{f,t} + u^{j,C}_{v,h,t}$$

Figures 4, 5, and 6 shows the estimated loadings $\hat{\beta}^{j,C}_{v,h,f}$ by variable (v), grouped by horizon (h), for forecaster type ($j = 10, 90$) for factor number f. The figures shows loadings for forecaster type 90 and 10 compared to the Consensus forecast loadings.

The top panel of each figure shows the loadings of the 90th type and the bottom panel the 10th type, both next to the loadings of the Consensus loadings. As can be seen in each figure, the Consensus loadings are in general of greater magnitude that the loadings by type. There appears to be shrinkage in the loadings of the forecast types relative to the SPF forecasts.

Forecasters have similar reduced-form models. The factors estimated from different forecasters are not only strongly correlated across forecaster, but within forecaster they display correlations with the same groups of variables. An additional point to note is that the loadings on the

Table 1: Canonical Correlations with SPF Consensus

Factor	10th	30th	50th	70th	90th
1	0.863	0.931	0.944	0.966	0.974
	(0.042)	(0.022)	(0.018)	(0.011)	(0.008)
2	0.848	0.902	0.920	0.929	0.952
	(0.046)	(0.030)	(0.025)	(0.023)	(0.015)
3	0.809	0.878	0.855	0.911	0.931
	(0.056)	(0.037)	(0.044)	(0.028)	(0.022)
4	0.748	0.857	0.842	0.890	0.895
	(0.071)	(0.043)	(0.047)	(0.034)	(0.032)
5	0.712	0.820	0.816	0.846	0.884
	(0.079)	(0.053)	(0.054)	(0.046)	(0.036)
6	0.684	0.792	0.750	0.787	0.856
	(0.085)	(0.061)	(0.071)	(0.062)	(0.043)
7	0.624	0.735	0.738	0.766	0.782
	(0.097)	(0.074)	(0.074)	(0.067)	(0.063)
8	0.558	0.692	0.709	0.751	0.738
	(0.108)	(0.084)	(0.080)	(0.071)	(0.074)
9	0.502	0.627	0.658	0.693	0.695
	(0.115)	(0.096)	(0.091)	(0.084)	(0.083)
10	0.406	0.569	0.626	0.651	0.652
	(0.124)	(0.106)	(0.097)	(0.092)	(0.092)

Canonical correlations correspond to the largest 10 eigenvalues of the matrix: $S_{CC}^{-1}S_{Cj}S_{jj}^{-1}S_{jC}$ where S_{CC}, S_{jj}, S_{Cj}, and S_{jC} correspond to the sample variance and covariance matrices of 20 factors estimated from the Consensus forecast C and percentile j.

Consensus forecasts are remarkably similar across forecaster type. Despite being from different forecasters, the underlying model of the forecasted data is remarkably similar across forecaster.

Figure 4: Shrinkage on Consensus Factor 1

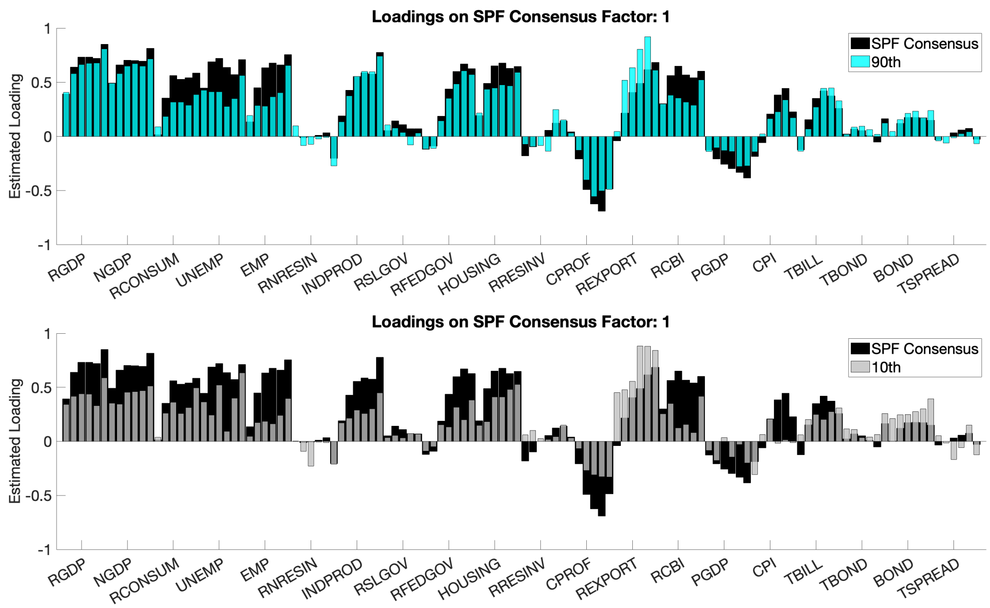


Figure shows the estimated coefficient, $\hat{\beta}^{j,C}_{v,h,1}$, from the regression: $e^{j}_{v,h,t} = \hat{\beta}^{j,C}_{v,h,1}\tilde{F}^{C}_{1,t} + u^{j,C}_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

Figure 5: Shrinkage on Consensus Factor 2

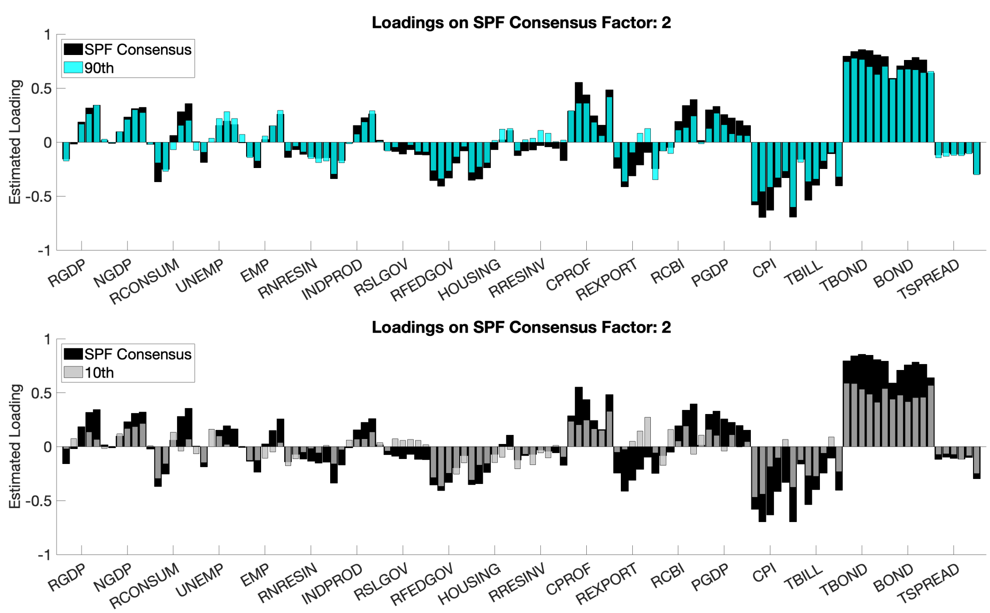


Figure shows the estimated coefficient, $\hat{\beta}^{j,C}_{v,h,2}$, from the regression: $e^{j}_{v,h,t} = \hat{\beta}^{j,C}_{v,h,2}\tilde{F}^{C}_{2,t} + u^{j,C}_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

Figure 6: Shrinkage on Consensus Factor 3

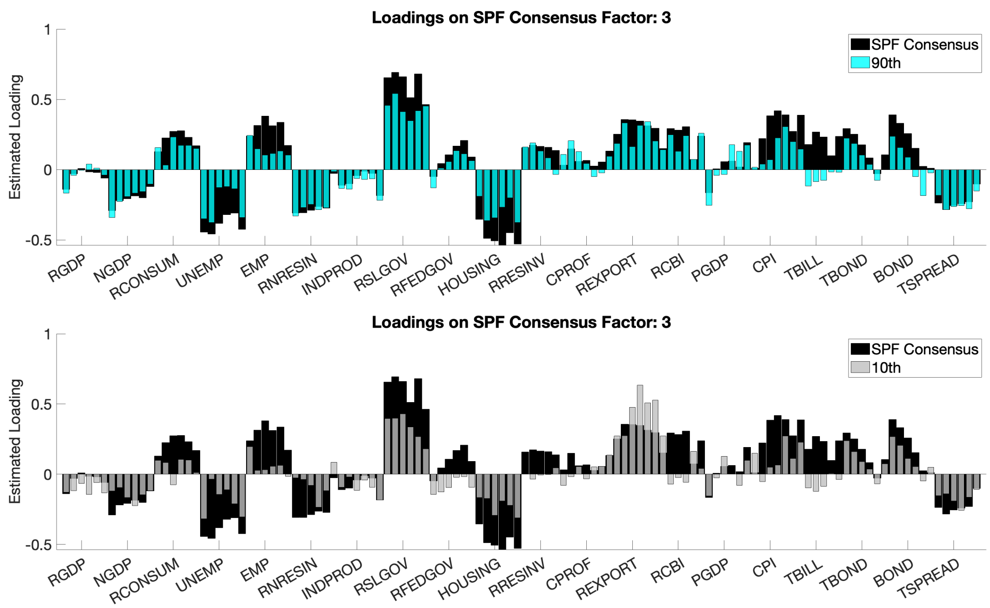


Figure shows the estimated coefficient, $\hat{\beta}^{j,C}_{v,h,3}$, from the regression: $e^{j}_{v,h,t} = \hat{\beta}^{j,C}_{v,h,3}\tilde{F}^{C}_{3,t} + u^{j,C}_{v,h,t}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$. Only one variable name corresponding to v is shown on the X-axis, variables are grouped together in order of forecast horizon.

2.4 Estimating Shrinkage Relative to the Consensus Forecast

For each type of forecaster, the plots in Figure 7 show 120 potential measures of shrinkage for each estimated factor. In this section, I propose a way to measure the shrinkage along each factor, so instead of 120 measures there would be a single scalar measure. Shrinkage can be captured by the diagonal matrix $\mathcal{S}$ situated between the factors and loadings estimated from a benchmark forecast:

$$\hat{X}^{j} = \underbrace{\tilde{F}^{C}}_{T\times k}\underbrace{\mathcal{S}}_{k\times k}\underbrace{\tilde{\Lambda}^{C\prime}}_{k\times N} + e^{j}$$

The elements of this matrix can be estimated by regressing the loadings of forecaster types 10 and 90 on the Consensus factors (the loadings estimated in the previous exercise) on the loadings of the Consensus forecasts on the Consensus factors. Section 4.3 describes the details of the estimation procedure and asymptotic distribution in more detail. Figure 7 shows the estimated shrinkage along the first ten factors for the two types of forecasters. The shrinkage is greater for the 10th percentile type, and is greater for higher order factors for both forecaster types.

Forecasters have the same reduced-form forecast model, but display shrinkage along the lower dimensions relative to the Consensus forecast. This shrinkage is increasing in the factor order - there is more shrinkage for less important factors. Additionally, econometric tests (Bai and Ng (2002), PC2) indicate 6 and 10 common factors in the 10th and 90th percentile forecasts,

Figure 7: Shrinkage along Consensus Factors

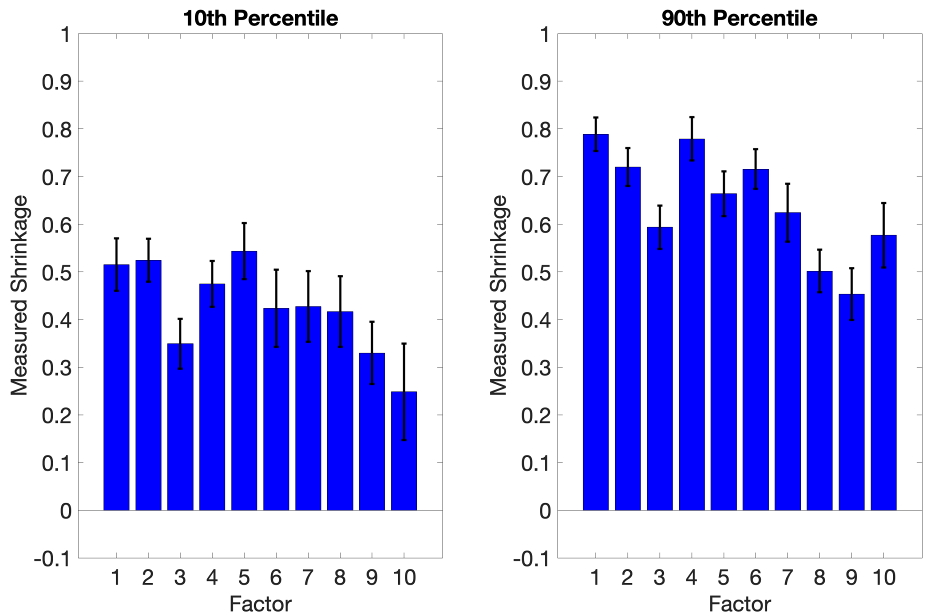


Figure shows the estimated coefficients, $\tilde{S}_{j,k}$, from the regression: $\hat{\beta}^{j,C}_{v,h,k} = \tilde{S}_{j,k}\hat{\beta}^{C,C}_{v,h,k} + \epsilon^{j}_{v,h,k}$ for $v = 1, \ldots, 20$ and $h = 0, 1, 2, 3, 4, \overline{2Y}$ for factors $k = 1, 2, \ldots, 10$.

respectively, and 20 common factors in the Consensus forecast.

The observed shrinkage patterns (more shrinkage for worse performers, more shrinkage for less important dimensions) are consistent with a rational inattention model. If the reason some forecasters are better is because they acquire better information about the underlying economic processes, we would expect to see less shrinkage relative to the benchmark consensus forecast. Additionally, if forecasters make a cost-to-benefit assessment along each dimension when acquiring information, they will acquire relatively less information about less important dimensions of information. The next section presents a theoretical model that can explain these empirical findings.

3 Rationally Inattentive Forecasts

In this section, I present a model of rationally inattentive forecasters which provides the basis for the mapping between rational inattention models and factor structure, and subsequently for my proposed estimator to recover attention. I consider a standard static linear-quadratic-Gaussian forecasting problem in which agents face a broad class of cost functions - separable spectral. Before describing the cost function in greater detail, I present the preliminaries of the agent's forecasting problem.

3.1 Preliminaries: A Rationally Inattentive Forecasting Problem

I begin by specifying the benefits of information for an agent making a forecast. Forecasters make predictions for n variables and have the goal of minimizing their squared forecast error. For example they could be forecasting GDP growth, inflation, and the change in interest rates. The true realization of the variables is denoted by X and the agents forecast is denoted $\hat{X}$. Before making a forecast, agents get to choose an information set $\mathcal{I}$ which is subject to a cost $C(\mathcal{I})$, to be further specified in the subsequent subsection. More information helps forecasters make more accurate forecasts. The agent's objective function is as follows:

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

The variables of interest linearly depend on a set of latent states, $\mathbf{a}$. 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.

$$\underbrace{X}_{n\times 1} = \underbrace{B'}_{n\times r}\underbrace{\mathbf{a}}_{r\times 1} + \underset{n\times 1}{u} \qquad\qquad r \leq n$$

The latent states ($\mathbf{a} = (a_1 \ldots a_r)'$) is potentially observable, whereas the $n \times 1$ vector u is unobservable. Both processes are Gaussian and independently and identically distributed:

$$\begin{aligned} &\mathbf{a} \sim N(0, \mathbb{I}) \\ &u_i \sim N(0, \sigma_{ui}) \text{ for } i = 1, \ldots, n \end{aligned}$$

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

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

Informationally-constrained forecasters differ from full information forecasters in that their forecasts are based on their belief about the observable component of the state instead of the true realization.

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

Assumption 1. *Forecasters have the same forecast model, characterized by the matrix B.*

Full information agents and rationally inattentive agents use the same model to make their forecasts. This model is characterized by the elements of the B matrix which maps the belief of the latent states to optimal forecasts. Prior beliefs about the variance of these states can be subsumed into B as long as forecasters share the same prior beliefs.

Assumption 2. *Forecasters have the same weight matrix, $W_j = W \ \forall j$.*

Forecasters value forecast performance in the same way. A forecaster might place more weight on getting their near-term forecast for GDP correct than their long-term forecast for GDP correct. Forecasters share these preferences. Without loss of generality the weight matrix W can be subsumed into the B matrix.

One way of thinking about assumption 1 is to suppose that all agents have have access to the history of realizations of the variables in question as well as precise estimates of the latent factors. With the same econometric model, they would come to the same coefficients on latent factors. Full information agents would track every availble signal about the innovation to these states, whereas rationally inattentive agents would only track some of them. Knowing that they acquired less than perfect signals, rationally inattentive agents would then hold different beliefs about the realization of latent states, reflecting their greater uncertainty.

Assumption 3. *Forecasters can only acquire Gaussian signals of* $\mathbf{a}$.

Agents form this belief through their acquisition of costly signals. They are restricted to choosing signals which are linear combinations of the latent states with Gaussian error:

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

In addition to the Gaussianity assumption, I further assume that the set of possible signals is rich.

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

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

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

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

While the agents can choose both the signal structure and signal precision, the minimization goal in the agent's objective function can be written exclusively as a choice of the expected posterior variance.

$$
\begin{aligned}
&\mathbb{E}\left[(\hat{X}-X)'(\hat{X}-X)\right] \\
&=\mathbb{E}\left[(B'\hat{\mathbf{a}}-B'\mathbf{a}-u_i)'(B'\hat{\mathbf{a}}-B'\mathbf{a}-u_i)\right] \\
&=\text{tr}\left(BB'\hat{P}\right)+\sum_{i=1}^{n}\sigma_{ui}
\end{aligned}
$$

The first part of the objective function can be written as a choice of expected posterior variance. The second part of the equation, reflecing the variance of the unobservable component, will not affect the optimization problem and so is dropped from future expressions.

I next turn to the specification of the cost function and consider a class of functions which are defined over the expected divergence between prior and posterior variance beliefs. This class of functions has desirable optimization characteristics, and also subsumes commonly-used information costs in the economics literature.

3.2 Separable Spectral Cost Functions

The measure of attention proposed in this paper is valid for a broad and novel set of cost functions. This section introduces that class - separable spectral cost functions. These cost functions are defined as the difference between additively separable functions of the eigenvalues of the prior and posterior variance matrices. The proposed measure of attention captures the shrinkage of eigenvalues in the rationally inattentive forecast relative to the best possible forecast, and so has an intuitive connection to cost functions defined over eigenvalues.

Rational inattention models, and models with endogenous information choice more broadly, require a characterization of the structure of the agent's information choice, as well as how to define a cost over it. Intuitively, an agent has acquired information if their uncertainty about a unknown entity declines. Uncertainty is intimately related to the probability distributions of random events. An agent would be less uncertain about the outcome of a variable with a very low variance than very high variance - it rains so rarely in Phoenix that looking at the weather forecast would do little to change one's uncertainty. Information acquisition can be captured by changes in the distribution of an agent's belief about the outcome of a random event.

Rational inattention models include a choice of information given a cost on acquisition, and the specification of the cost function can lead to different predictions about optimal attention allocation. Sims' (2003) initial work on rational inattention theory focused on a constraint characterized by mutual information, or a reduction in Shannon entropy which is popular in the signal processing literature (Cover and Thomas (2006)). While mutual information has proven useful for its tractability, its predictions are contradicted by observed experimental evidence (Dean and Neligh (Forthcoming)). Hébert and Woodford (2021) consider a neighborhood based cost function which has the empirically-grounded feature that agents have an easier time distinguishing states that are more different. In the context of a continuous linear-quadratic Gaussian tracking problem, neighborhood-based cost functions simplify to an average of the state space over the Fisher information, a tractable alternative to Mutual Information. More generally, Caplin et al. (2022) study large class of functions they define as uniformly posterior separable cost functions. Under UPS cost functions the prior and posterior variance matrices are assumed to be additively separable. UPS cost functions can arise from sequential accumulation of information. Mutual information and the empirically more realistic Fisher information are subsumed under UPS cost functions.

In rational inattention problems agents make a choice over their expected distribution of posterior beliefs, given their cost function. This distribution is characterized by a signal structure. If an agent chooses a very precise signal, the potential posterior beliefs will tightly distributed.

The model considered in this paper is restricted to Gaussian shocks and signals, implying that the set of potential posterior distributions will also be Gaussian. The posteriors can therefore be fully parameterized by the mean and variance. Additionally, given a specific signal choice (e.g. a choice of Z and H), the agent's posterior beliefs will only differ in the mean. Thus, by defining a cost function exclusively over a divergence between prior and posterior variance, one can ignore the expectation over distributions of posteriors.

The eigenvalues of a variance matrix capture the degree of variability, and hence uncertainty, along the dimensions of the set of random variables. Higher relative eigenvalues correspond with higher relative variance shares. For illustration, consider a 3×1 vector of unbiased signals, $\tilde{\mathbf{x}}$, for a variable of interest $\mathbf{x}$:

$$\begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \tilde{x}_3 \end{bmatrix} \sim N\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \Sigma \right) \qquad \Sigma = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix}$$

The diagonal elements of the variance matrix are the eigenvalues. Suppose $\sigma_1 \gg \sigma_2,\ \sigma_3$. Most of the noise in the signal of $\mathbf{x}$ would come from the first element of the vector. A better signal for x_1 would be captured by a lower variance in the signal, $\hat{\sigma}_1 < \sigma_1$, and consequently, a lower eigenvalue. Costs defined over eigenvalues capture this relationship between the eigenvalues and the precision of the signal.

For a general covariance matrix (not necessarily diagonal), the eigenvalues describe the relative importance of dimensions along the eigenvectors of the covariance matrix. Separable spectral cost functions are defined over the reduction in uncertainty along dimensions of the eigenvectors of the variance matrix.

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

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

Additionally:

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

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

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

The separable spectral cost function subsumes known cost functions. Both the trace and determinant operators are defined over eigenvalues, and both feature prominently in current cost functions. Under a Gaussian latent process and signal structure, the cost function defined by mutual information (MI) is characterized as the difference between the log of the determinant of the posterior and prior variance matrices. This amounts to differences in the sums of the log of the eigenvalues of each matrix - a cost defined over the eigenvalues. Let d_i be the i^{th} eigenvalue of $\hat{P}$, the function $\mathbf{f}$ in the case of mutual information is then:

$$\mathbf{f}^{MI}(\hat{P}) = -\sum_{i=1}^{m} \log\left(d_i\right)$$

Another popular type of cost function is characterized by the trace of the precision matrix (TP) (Kacperczyk et al. (2016), Hébert and Woodford (2021)). The trace of a matrix is equivalent to a sum of the eigenvalues, and the eigenvalues of a positive definite matrix are simply inverted in the inverse matrix operation. The cost function can again be defined as a cost function over eigenvalues:

$$\mathbf{f}^{TP}(\hat{P}) = \sum_{i=1}^{m} d_i^{-1}$$

3.2.1 Optimal Choice of Posterior Variance

In this section I return to the agent's objective function. As previously discussed, both the objective function and cost can be written in terms of the posterior belief about the variance of the latent factors, $\bar{P}$, given the choice of signal structure. Prior variance beliefs are normalized to $\mathbb{I}$. The agents also face the constraint that $P_t - P_{t|t} \succeq 0$ - what is typically referred to as the "no-forgetting" constraint. Agents enter the forecasting period with a prior belief about the variance of the latent states. This might be a more precise signal that would be optimal for that dimension of the information space, but they cannot lose precision along that dimension to add precision along another dimension without paying an additional cost.

$$\min_{\{\hat{P}\}} \text{tr}\left(BB'\hat{P}\right) + \mathbf{f}\left(\hat{P}\right)$$
$$\text{s.t. } \mathbb{I} - \hat{P} \succeq 0$$

The spectral decomposition of the posterior variance matrix is $\bar{P} = UDU'$, where d_i denotes the i^{th} diagonal element of D, and u_i the i^{th} column of U. Using the fact that $\text{tr}\left(BB'\bar{P}\right) = \sum_{i=1}^{m} d_i u_i' BB' u_i$, I rewrite the objective function in terms of a choice of eigvalues and eigenvectors of the posterior variance[7]:

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

Lemma 1. *Given a separable spectral cost function, the agent's forecasting problem can be written as separate choices of both the eigenvectors of the posterior variance ($\mathbf{u} = (u_1, \dots u_p)$) and eigenvalues ($\mathbf{d} = (d_1, \dots d_p)$).*

Lemma 1 captures the rotational invariance feature of separable spectral cost functions. Costs are only defined over the eigenvalues of the posterior variance, or the precision of forecast along the dimension considered, but not the dimension itself.

$$\min_{\{d_i\}_{i=1}^p} \sum_{i=1}^{p} d_i \min_{\substack{\{u_i\}_{i=1}^p \\ u_i \perp u_1, \dots u_{i-1}}} \left[u_i' BB' u_i\right] + \omega \sum_{i=1}^{p} f\left(d_i\right)$$
$$\text{s.t. } 0 \le d_i \le 1$$

Consider the term $u_i' BB' u_i$ in the first part of the problem. As u_i is an eigenvector, $u_i' u_i = 1$,

[7] I ignore the prior variance in the objective function, as it will not affect the optimal attention allocation.

thus allowing for use of the Raleigh-Ritz and Courant-Fisher-Weyl min-max principals[8]. Define pairs of eigenvectors and eigenvalues from the spectral decomposiiton as $U^B = \left(u_1^B, u_2^B, \dots u_p^B\right)$ and $\Lambda^B = \left(\lambda_1^B, \lambda_2^B, \dots \lambda_p^B\right)$ with eigenvalues in descending order ($\lambda_1^B \geq \lambda_2^B \geq \cdots \geq \lambda_p^B$). From Raleigh-Ritz, for $i = 1$:

$$\begin{aligned} u_1^B &= \text{argmin}\left[u_1' BB' u_1\right] \\ \lambda_1^B &= \left[u_1^{B\prime} BB' u_1^B\right] \end{aligned}$$

From Courant-Fisher-Weyl , for $i \geq 2$:

$$\begin{aligned} \lambda_{m-i+1}^B &= \underbrace{\text{min}}_{u_i \perp u_1, \dots u_{i-1}} \left[u_i' BB' u_i\right] \\ u_i^B &= \text{argmin}\left[u_i' BB' u_i\right] \end{aligned}$$

The minimum of the optimization of the term $u_i' BB' u_i$ occurs along the eigenvectors of BB' and is equivalent to the eigenvalues. With this characterization, the agents optimization problem can be written exclusively as a choice of eigenvalues of the target posterior variance matrix. Eigenvectors of that matrix are given by U^B, which captures the optimal dimensions along which to acquire information. The choice of eigenvalues is described in Theorem 1.

Proposition 3.1. *Let $\lambda_1^B, \dots \lambda_p^B$ denote the eigenvalues of BB' ordered from largest to smallest. The informationally-constrained forecasters objective function can be written in terms of a choice of eigenvalues of the posterior variance matrix:*

$$\begin{aligned} \underset{\{d_i\}_{i=1}^p}{\mathit{min}} \sum_{i=1}^p \lambda_i^B d_i + \omega \sum_{i=1}^p f\left(d_i\right) \\ \mathit{s.t.}\ 0 \leq d_i \leq 1 \end{aligned}$$

An agent will target a posterior variance with eigenvalues, d_i, that satisfy $\lambda_i^B = -f'\left(d_i\right)$, with the

[8]**Courant-Fisher-Weyl Min-Max Principal.** Let A be a positve definite symmetric $n \times n$ matrix with eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \dots \lambda_n(A)$ and corresponding eigenvectors $u_1, u_2, \dots, u_k \in \mathbb{R}^n$. Let k with $1 \leq k \leq n$ and let $x \in \mathbb{R}^n$ and $x \neq 0$. Then

$$\lambda_k(A) = \underbrace{\text{min}}_{u_1, u_2, \dots u_{k-1}} \underbrace{\text{max}}_{x \perp u_1, u_2, \dots u_{k-1}} \frac{x' A x}{x' x}$$

and

$$\lambda_k(A) = \underbrace{\text{max}}_{u_1, u_2, \dots u_{n-k}} \underbrace{\text{min}}_{x \perp u_1, u_2, \dots u_{n-k}} \frac{x' A x}{x' x}$$

constraint that $0 \leq d_i \leq 1$.

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

First order conditions imply that the cost to benefit ratio $\frac{\lambda_i^B}{\omega} = f'(d_i)$ will be satisfied. If the marginal cost of paying attention to the i^{th} dimension (characterized by $f'(d_i)$) exceeds the prior variance (1), then the optimal choice of d_i will be 1 (agents will pay no attention). Agents will only acquire signals about dimensions of the information space, and thus have a change from prior to posterior variance belief along that dimension of the information space, when the benefit of the signal exceeds the cost. The targeted posterior variance can be written as:

$$\bar{P}^* = U^B \Omega U^{B\prime}$$
$$\Omega(i,i) = \min\left(1, f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right)$$

What does the solution to the agent's forecasting problem imply? Agents target a specific posterior variance with their choice of signals, taking into consideration the relative importance of dimensions of the posterior. The eigenvalues of the matrix BB' reflect the relative importance of the underlying latent factors $\mathbf{a}_t$ to the forecasted variables. An agent will pay more attention to a factor that is more important for more of the forecasted variables. The thresholding follows from the no-forgetting constraint. The implied thresholding patterns will be determined by the elements of Ω.

It remains to be shown that for a given choice of optimal posterior variance, there exists a choice of signal structure (Z, H) that leads to the optimal posterior variance.

Lemma 2. *There exists a signal structure* (Z, H) *that can achieve the agent's optimal posterior variance.*

The proof is given in the appendix.

3.3 Rational Inattention with Dynamic Factors

In this subsection, I generalize the cost function to allow for dynamic factors. I will argue that the intuition from the static model extends to the dynamic case under a realistic assumption. The agent's objective function now includes a time subscript.

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

The mapping from state to optimal forecast remains the same. So long as forecasters share the same weight matrix, the mapping characterized by B will remain the same.

$$\hat{X}^* = B'\mathbf{a}_t$$

In the dynamic case the latent states $\mathbf{a}_t$ have the following dynamic process:

$$\underbrace{\mathbf{a}_t}_{p\times 1} = \sum_{k=1}^{q} \underbrace{\mathcal{T}_k}_{p\times p} \underbrace{\mathbf{a}_{t-k}}_{p\times 1} + \underbrace{\eta_t}_{p\times 1} \qquad \eta_t \sim N(0, Q)$$

The dynamic problem has distinct predictions from the static problem, as information will have more value for more persistent processes[9].

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 4. *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 $\mathbf{a}_t$.*

As prior signals took the form $s_t = Z\mathbf{a}_t + \epsilon_t$, the perfect signal of a_t would subsume any information previously acquired. The prior belief in the variance of is again unity, which is equivalent to the following restriction:

Assumption 5. $\sum_{k=1}^{q} \mathcal{T}_k \mathcal{T}_k' + Q = \mathbb{I}$. *Var* $(\mathbf{a}_t) = \mathbb{I}$.

With these additional assumptions, the intuition from the static problem carries through to the dynamic case.

[9] Afrouzi and Yang (2021) explore the resulting changes in attention allocation in this case.

3.3.1 Agent Belief Updating with Dynamic Factors

Agents again update their beliefs of the state after receiving signals about the underlying state. Given two additional assumptions, the agent's posterior variance remains $\hat{P}$:

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

The updated belief about the mean, however, takes a different form:

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

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

In the dynamic case, the agent's belief about the state, $\hat{\mathbf{a}}_t$, reflects the known persistence, characterized by the coefficients of $\mathcal{T}$. Any change from this baseline autoregressive belief will reflect signals acquired about the innovation away from that process.

The autoregressive coefficient of GDP growth on its lag between 1984 and 2019 is 0.74, indicating a high degree of persistence. An analyst making a forecast of GDP growth could achieve a reasonable forecast just by taking this persistence into account. For a better forecast, however, they might pay attention to the various monthly data releases that could shed light on current quarter innovations away from their expectation.

In this way, the rationally inattentive forecaster's 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 observation equation error variance (H) 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 purchase weekly credit and debit card spending from Affinity to get a higher-frequency signal on consumer spending, or job posting data from Lightcast to get an additional signal on underlying economic growth. These additional signals come at an additional cost to the analyst.

3.3.2 Attention in Forecasted Variables

Given their updated belief about the state vector, agents make the following forecast:

$$
\begin{aligned}
X_t &= B'\hat{\mathbf{a}}_t \\
&= B'\sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k} + B'\left(\mathbb{I} - U^B\Omega U^{B\prime}\right)\left(\mathbf{a}_t - \sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k}\right) + B'\xi_t \\
&= \underbrace{B'\sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k}}_{\text{Autoregressive}} + V^B D^B \underbrace{(\mathbb{I}-\Omega)}_{\text{"Attention" Matrix}} U^{B\prime}\underbrace{\left(\mathbf{a}_t - \sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k}\right)}_{\text{Signal Surprise}} + B'\xi_t
\end{aligned}
$$

The implied attention of agents is the diagonal matrix $(\mathbb{I}-\Omega)$ which characterizes the degree to which agents update their prior beliefs about the state. In the static case, the prior belief was 0. In the dynamic case, the prior belief reflects the previous observation of latent states

$$
(\mathbb{I}-\Omega) =
$$

$$
\begin{pmatrix}
\max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_1^B}{\omega}\right)\right) & 0 & \dots & 0 \\
0 & \max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_2^B}{\omega}\right)\right) & \dots & 0 \\
0 & 0 & \ddots & \vdots \\
0 & 0 & \dots & \max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_p^B}{\omega}\right)\right)
\end{pmatrix}
$$

The error term is defined as follows: $\xi_t = Z'\left(ZZ' + H\right)^{-1}\epsilon_t$. Agent's update their belief about the underlying state according to the multiplier $\left(\mathbb{I} - U^B\Omega U^{B\prime}\right)$ on the difference between their belief of the state, and that indicated by the signals they receive. Compare this to a full information agent's forecast, in which the forecast does not include the extra attention matrix:

$$
\begin{aligned}
X_t^* &= B'\mathbf{a}_t \\
&= B'\sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k} + V^B D^B U^{B\prime}\left(\mathbf{a}_t - \sum_{k=1}^{q}\mathcal{T}_k\mathbf{a}_{t-k}\right)
\end{aligned}
$$

The matrix Ω, with elements between 0 and 1, captures the attention of a rationally inattentive agent to a dimension of the information space relative to the full information agent. Rationally inattentive forecasts load less on the realized state variables, reflecting agent information costs.

3.3.3 Information at Different Horizons

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 housing starts ($x^h_{S,t},\ x^h_{L,t}$). There are 4 latent factors which correspond to: short-term growth, short-term housing orthogonal to growth, long-term growth orthogonal to short-term factors, and long-term housing orthogonal to everything else.

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

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

Short-term housing starts load both on the short-term growth and short-term housing-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 housing 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 commodity-traded price of lumber today could eventually affect the decision to build a new house through the cost channel, but it would unlikely affect near-term housing starts. 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.

4 Rational Inattention through the Lens of a Factor Model

In this section I present the manifestation of information constraints in the principal components of the theoretically generated forecast data. A panel of rationally inattentive forecasts theoretically will include an extra shrinkage and thresholding "attention" matrix on the eigenvalues as compared to the full information forecast. Rationally inattentive agents are in effect performing singular value thresholding on the information space relative to their full information peers, resulting in different predicted factor structures.

From this section onwards, rationally inattentive forecasters are indexed by their rank j, which *only* affects their cost of information. The rank j corresponds to the performance ranking, e.g. $j = 20$ corresponds to the 20th percentile forecaster. The j^{th} type forecaster solves the following problem:

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

Forecasters *only* differ in their marginal cost of information acquisition, ω_j. This will affect their choice of information acquisition, $\mathcal{I}_t^j$, as well as their forecast, $\hat{X}_t^j$.

The forecast vectors at time t are $\hat{X}_t^j$ and $\hat{X}_t^*$, which are of dimension $n \times 1$. The corresponding matrix forms are dimension $T \times n$ for the rationally inattentive, $\hat{X}^j$, and full information agents, $\hat{X}^*$, respectively. The theoretical model implies that agents only differ on the degree of attention to innovations outside of the autoregressive part of forecasts. Matrix forecasts $\tilde{X}^j$ and $\tilde{X}^*$ are forecasted variables that have been orthogonalized to the lagged values of the state, demeaned and normalized by their standard error and by the factor $\sqrt{NT}$[10].

$$\hat{X}_t^j = \hat{B}' \sum_{k=1}^{q} \mathcal{T}_k \mathbf{a}_{t-k} + \tilde{X}_t^j$$

$\tilde{X}_t^*$ is similarly defined. The rationally inattentive forecast matrix can be written as follows:

$$\begin{aligned}
\underbrace{\tilde{X}^*}_{T\times n} &= \underbrace{\mathbf{a}}_{T\times r} \underbrace{U^B}_{r\times r} \underbrace{D^B}_{r\times r} \underbrace{V^{B\prime}}_{r\times n} \\
\underbrace{\tilde{X}^j}_{T\times n} &= \underbrace{\mathbf{a}}_{T\times r} \underbrace{U^B}_{r\times r} \underbrace{D^B}_{r\times r} \underbrace{(\mathbb{I} - \Omega_j)}_{r\times r} \underbrace{V^{B\prime}}_{r\times n} + \underbrace{\xi^j}_{T\times r} \underbrace{B}_{r\times n}
\end{aligned}$$

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

[10] This normalization is consistent with the literature on principal component analysis, (e.g. Bai and Ng (2019)).

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 $\mathbf{U}^B \equiv \mathbf{a}U^B$ represents a new set of eigenvectors (the basis U^B extended along the time dimension).

$$\tilde{X}^j = \mathbf{U}^B D^B (\mathbb{I} - \Omega_j) V^{B\prime} + \underbrace{\xi^j B}_{e^j}$$

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

This representation makes it straightforward to pick out the r common factors and loadings from the best possible forecast:

$$\begin{aligned}\tilde{X}^* &= \mathbf{U_r}^B D_r^B V_r^{B\prime} \\ &= \underbrace{F^*}_{\mathbf{U_r}^B} \underbrace{\Lambda^*}_{D_r^B V_r^{B\prime}}\end{aligned}$$

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. Econometric procedures are then used to estimate the number of common factors.

Proposition 4.1. *Under assumptions 1-5, the rationally inattentive forecast implied by the model in section 3.3 can be written in terms of factors and loadings estimated from the full information forecast.*

$$\tilde{X}^j = F^* (\mathbb{I} - \Omega_j) \Lambda^{*\prime} + e^j$$

In summary, the informationally-constrained common component includes an extra shrinkage and thresholding matrix $(\mathbb{I} - \Omega_j)$ 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_j) that agents apply. The information choice of informationally-constrained agents has a number of interesting predictions.

4.1 An Estimator for Relative Attention

In this section I describe the procedure to recover the forecaster's attention matrix, $(\mathbb{I} - \Omega_j)$. While the theoretical model indicates that this should be a diagonal matrix, differences in the reduced form models of forecasters would lead to non-zero off-diagonal elements. For ease of notation, I will omit the j index for the derivation of the attention matrix. The estimator includes two steps: regress forecasted variables onto factors from the full information proxy forecast, and then regress resulting coefficients onto loadings from full information proxy forecast.

In the case that the forecaster allocates attention along different dimensions than the best possible forecaster, or has a different reduced form model, the resulting attention matrix would not be diagonal. Let H be a roatation matrix, such that FH is the dimension along which rationally inattentive forecasters allocate attention. The estimated attention matrix would include this rotation.

$$\hat{X} = F^* H \left(\mathbb{I} - \Omega\right) H^{-1} \Lambda^{*\prime} + e$$

Define $\mathit{\Phi}$ as the rotated attention matrix, where $\mathit{\Phi}_{x,y}$ denotes the x^{th} row and y^{th} column.

$$\mathit{\Phi} = H \left(\mathbb{I} - \Omega\right) H^{-1}$$

The rationally inattentive forecast can now be written as:

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

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:

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

Assumption 6. *For factor $k \in 1, \dots, r$*

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

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 $i \in 1, \dots, N$ regress $T \times 1$ vector $\hat{X}_i$ on $T \times r$ matrix $\tilde{F}^*$. This results in N estimates for the k^{th} ($k \in 1, \dots r$) coefficient, $\overline{(\lambda^*_{1i}\Phi_{k,1} + \lambda^*_{2i}\Phi_{k,2} + \dots + \lambda^*_{ri}\Phi_{k,r})}$.

2. For each $k \in 1, \dots, r$ 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, $(\widehat{\Phi_{k,1}}, \widehat{\Phi_{k,2}}, \dots \widehat{\Phi_{k,r}})$.

The asymptotic distribution of the estimator is summarized in Proposition 4.2.

Proposition 4.2. *Suppose Assumption 6 holds. Under the normalization $\frac{F'F}{T} = \mathbb{I}$, the proposed estimator for the k^{th} row ($k \in 1, \dots, r$) of the rotated attention matrix Φ converges to the following distribution.*

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

Where $\Sigma_{\Lambda e F_k}$ is given in Assumption 6, and $\Sigma_\Lambda = \frac{\Lambda'\Lambda}{N}$.

Figures in the appendix show the distributions of estimated attention from monte carlo simulations, which demonstrate the consistency of the estimator.

4.2 Consensus Forecast as Theoretical Best Possible

A forecast comprised of the median forecasted value for each variable, or the Consensus forecast, will characterize the optimal basis for a forecast. Additionally, under certain conditions, it will be a superior forecast than any forecasted value by an individual. I characterize these conditions. Let b_i denote the i^{th} column of B. The forecast for variable i from forecaster type j can be written as follows:

$$\begin{aligned} X_{it}^{j} &= b_i' \left(\mathbb{I} - P_j\right) \mathbf{a}_t + b_i' Z_j' \left(Z_j Z_j' + H_j\right)^{-1} \epsilon_t^j \\ \text{where } P_j &= \mathbb{I} - Z_j' \left(Z_j Z_j' + H_j\right)^{-1} Z_j \end{aligned}$$

Both $\mathbf{a}$ and ϵ_{jt} are normally distributed, thus the distribution of X_{it}^j will also be normally distributed. Under a normal distribution, the median and mean will be the same. To get the mean, I take the expectation over forecaster type. Let $\mathbb{E}\left[\cdot\right]_j$ denote the expectation over forecaster type j.

$$\begin{aligned} \mathbb{E}\left[X_{it}^j\right]_j &= b_i' \left(\mathbb{I} - \mathbb{E}\left[P_j\right]_j\right) \mathbf{a}_t + b_i' Z_j' \left(Z_j Z_j' + H_j\right)^{-1} \mathbb{E}\left[\epsilon_t^j\right]_j \\ &= b_i' \left(\mathbb{I} - \mathbb{E}\left[P_j\right]_j\right) \mathbf{a}_t \end{aligned}$$

The performance of each forecast is the magnitude difference from the best possible (BP) forecast ($X_{it}^{BP} = b_i' \mathbf{a}_t$).

$$\begin{aligned} |\mathbb{E}\left[X_{it}^j\right]_j - X^{BP}| &= b_i' \mathbb{E}\left[P_j\right]_j \mathbf{a}_t \\ |X_{it}^j - X^{BP}| &= b_i' P_j \mathbf{a}_t + |b_i' Z_j' \left(Z_j Z_j' + H_j\right)^{-1} \epsilon_t^j| \end{aligned}$$

Lemma 3. *If the following inequality holds:*

$$\mathbb{E}\left[P_j\right]_j \mathbb{E}\left[P_j\right]_j' - P_j P_j' < Z_j' \left(Z_j Z_j' + H_j\right)^{-1} H_j \left(Z_j Z_j' + H_j\right)^{-1} Z_j$$

The median forecast will have a lower forecast error than forecaster type j.

Lemma 3 captures the tradeoff between attention and noise. The median forecast will, in the limit, not have noise in the observed signal, but could feasibly display shrinkage along the factors relative to a forecaster with very low cost. Estimated values of the cost and benefits of acquiring information in Section 5 indicate that this inequality is always satisfied in the macroeconomic forecast data set considered. Indeed, the consensus forecast performs better than the high-type forecasters (90th percentile) 90% of the time.

The SPF consensus forecast additionally will have an unbiased basis relative to the theoretical best possible. The basis is captured by the the rotation of shocks $\mathbf{a}_t$ by the vector b_i. Even top performing forecasters will be shifted away from this basis. Shrinkage, or relative attention, is captured along the basis of forecasts, making that the SPF consensus is an ideal benchmark.

4.3 Relative Attention to Identified Dimensions

The previous section provided a framework to measure relative attention by agents to different eigenvectors of the information space. This section covers the necessary assumptions and restrictions on forecasting frameworks that permit an economic interpretation of the factors, itself valid for researchers.

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 identification 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[11].

Lemma 4. *Let the coefficient matrix B satisfy the following:*

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

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

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.

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

[11]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

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

$$\tilde{F}^* = \hat{F} \cdot Q \qquad\qquad \tilde{\Lambda}^* = \hat{\Lambda}^* \cdot Q = \begin{pmatrix} R' \\ \hat{\Lambda}_2^* \end{pmatrix}$$

Attention can then be measured along the new basis of factors.

5 Empirical Application

In this section I return to data from the *Survey of Professional Forecasters* (SPF). I first revisit the observed shrinkage documented in Section 2. Through the lens of a rational inattention model, these estimated shrinkage can be interpreted as the attention of forecasters relative to the Consensus forecast. I next examine the claim that forecasters have the same reduced form model by looking at the estimated off-diagonal elements of the attention matrix and find only minor estimated differences in the higher-order factors, implying that forecasters use largely the same model. Returning to the theoretical model, I estimate the cost of information for forecasters of differing accuracy using different functional forms of the cost function, before estimating a parametric form of the cost function. Consistent with the first picture, I estimate higher cost values for lower-performing type forecasters. Finally, I consider attention to economically-meaningful factors and show that forecasters pay even less attention to factors associated with long-term forecasts of inflation.

5.1 Revisiting Shrinkage in Macroeconomic Forecasts

5.1.1 Macroeconomic Forecast Data

I return to data from the *Survey of Professional Forecasters* (SPF), and again consider forecasts of: 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 (3Mo), Treasury Bonds (10Y), Treasury Spreads (10Y - 3Mo), and BAA corporate bond spreads over 10Y Treasuries.

I look at forecasts for either the percent change, difference (unemployment, nonfarm payrolls, interest rates), normalized difference (inventories and exports), or levels (interest rate spread data) for the current quarter, and the subsequent four quarters, as well as the two-year ahead annual forecast[12]. Additionally, to remove the autoregressive component of the forecasts, I orthogonalize forecasted variable to one lag of the real-time observations[13] of the variable ($N = 120$, $T = 144$).

As forecasters drop in and out of sample, I construct forecast panels by type which correspond to the forecaster's historic annual average forecast performance. The forecasted vector at time t for rank j comes from the forecaster who ranked at the p^{th} percentile[14] in the forecast made at $t-4$ quarters ago (excluding at the two-year ahead annual forecasts).

As a proxy for full information I consider both the SPF Consensus Forecast and the Federal Reserve Greenbook forecasts. The SPF Consensus forecast is the median forecasted value of each variable. The full forecast does not correspond to any individual forecaster. The Greenbook forecasts are made by Federal Reserve staff in advance of FOMC meetings and are released with a 5-year lag. The Greenbook forecast does not include inventories, exports, employment, corporate profits, BAA corporate bond spreads, Treasury spreads and does not include a 2-year ahead annual forecast for any variable ($N = 70$, $T = 136$).

5.1.2 Shrinkage as Relative Attention

In Section 2, I documented that professional forecasts displayed shrinkage relative to the Consensus benchmark and that lower-performing forecast types displayed more shrinkage. Additionally, the degree of shrinkage was shrinking along factor order.

Figures 8 and 9 shows the measured shrinkage of professional forecasters relative to the SPF Consensus benchmark and the Federal Reserve Greenbook benchmark. Under the rational inat-

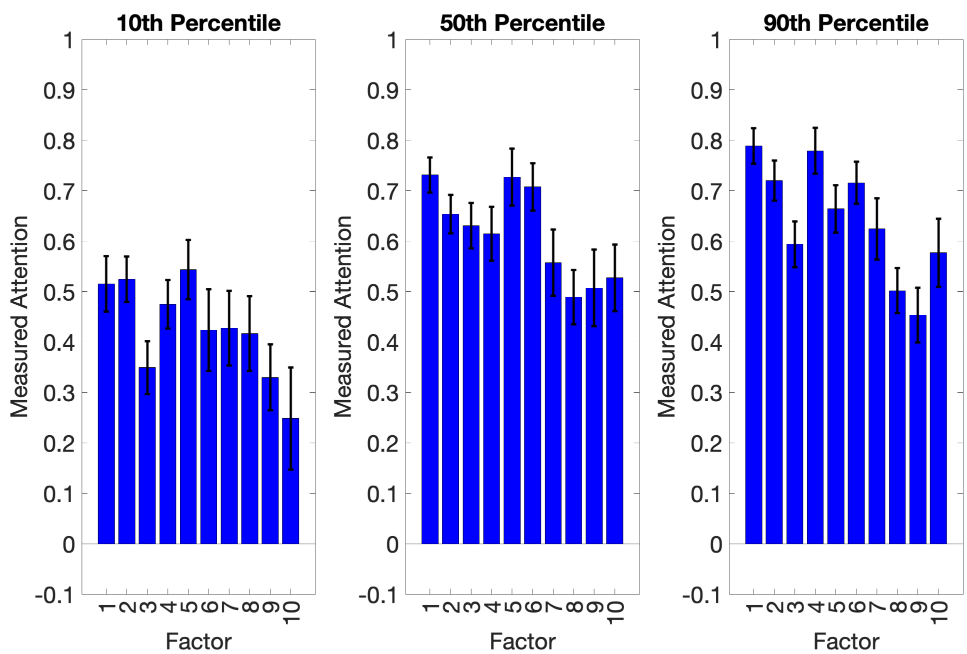
[12] Annual forecasts for the next year are available, but are excluded so as not to overlap with the quarterly forecasts.

[13] The two-year ahead annual forecasts are orthogonalized to the mean of the past four quarters of growth (annualized).

[14] Percentile is in terms of performance, 1^{st} percentile would be worst, and 100^{th} percentile best.

tention model, this shrinkage can be interpreted as relative attention. The general pattern is lower attention to higher order factors, and greater attention for higher types.

Figure 8: Relative Attention to SPF Consensus



5.1.3 Estimated Thresholding

Figure 10 displays the eigenvalues estimated by forecast panel. The lower-type forecasters in general have smaller eigenvalues, and fewer estimated factors using a factor selection criteria (Bai and Ng 2002, PC2 criteria). The criteria selects 6, 7, and 10 factors for the 10th, 50th and 90th percentile types, respectively. These are far fewer than the 20 selected for the SPF Consensus forecast. The test indicates that the lower-type forecasts include fewer common factors.

5.2 Examining Differences in Reduced Form Models

A forecaster's reduced form model is characterized by the eigenvectors of the variance matrix. If a forecaster uses the same model as the benchmark, the eigenvectors or basis of dimensions which

Figure 9: Relative Attention to Federal Reserve Greenbook

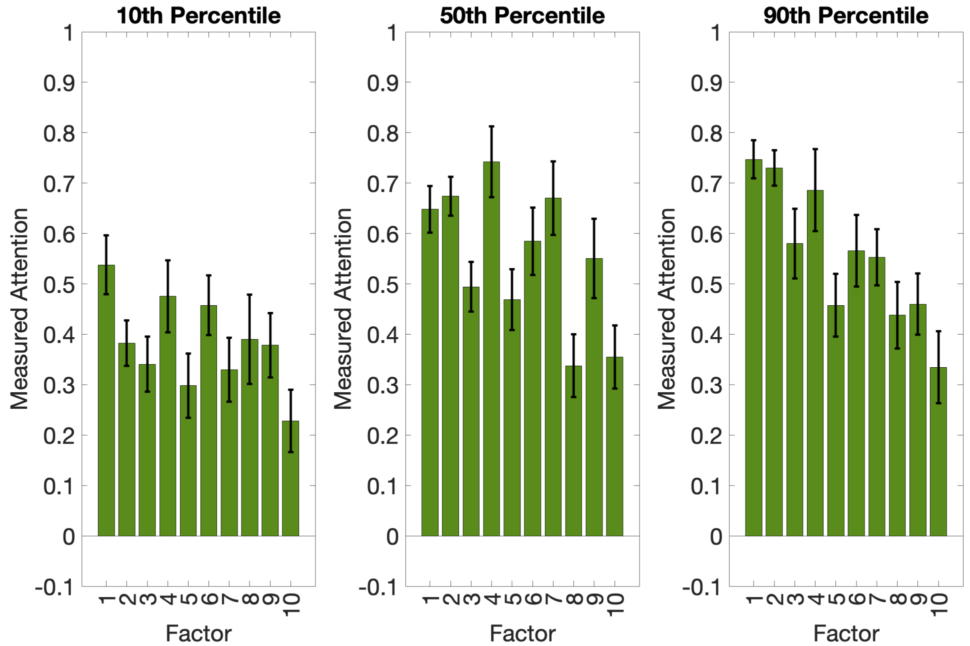
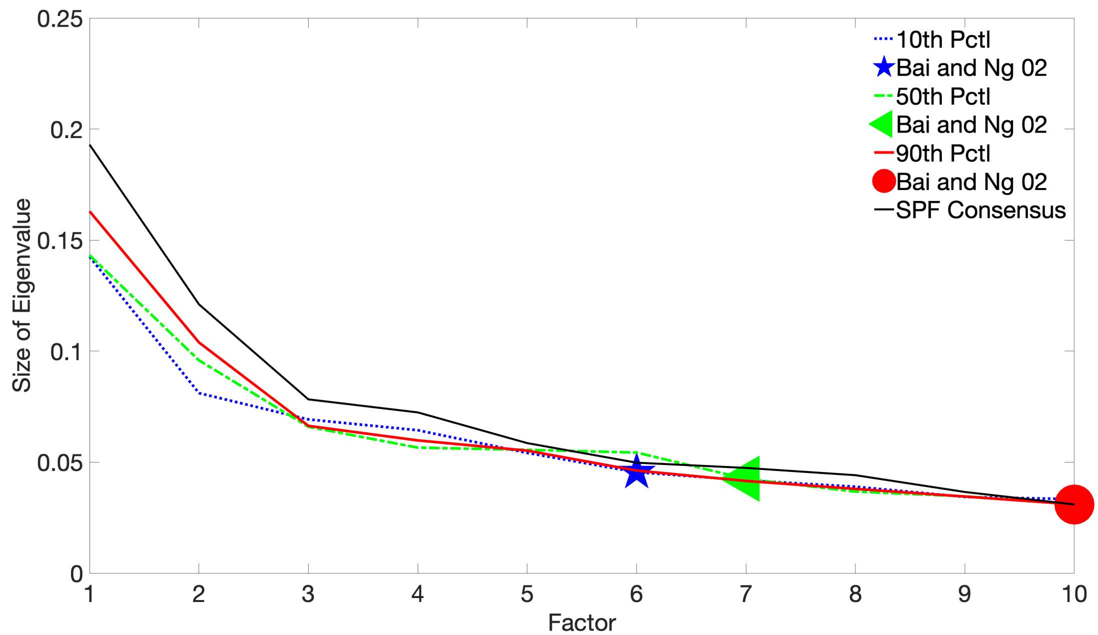


Figure 10: Size of Eigenvalues by Forecaster Type



Markers on plot indicate number of factors selected using the Bai and Ng (2002) PC2 criteria.

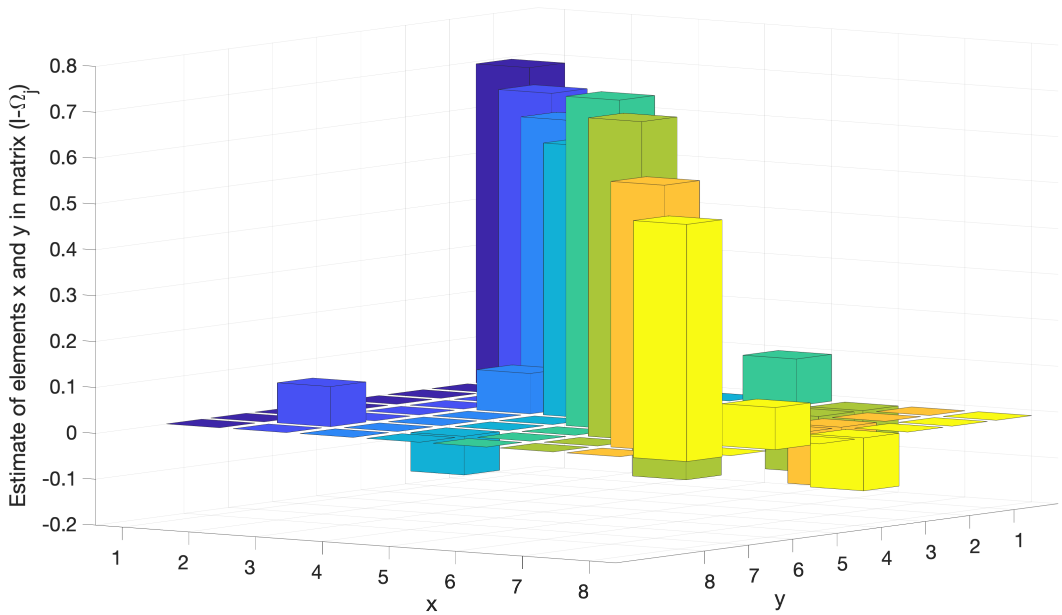
explain the variance in their forecast will be the same as those of the benchmark. If, however, a forecaster uses a different model, the eigenvectors of their variance matrix will be rotated away from the eigenvectors of the benchmark by some orthogonal matrix H.

The forecast from someone with a different reduced form model relative to the benchmark will include this extra H matrix:

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

The matrix $H\left(\mathbb{I} - \Omega\right) H'$ is not a diagonal matrix. Using the estimation procedure described in Section 4, I calculate the share of significantly non-zero off-diagonal elements. These estimated matrices are generally sparse, with the number of non-zero off-diagonal elements ranging from less than 4% to just over 20%. In magnitude, these elements are far less than the measured attention, as demonstrated in Figure 11.

Figure 11: Estimated Size of Elements of $\left(\mathbb{I} - \Omega_{50}\right)$



Point estimates from elements that are not significantly different from zero are set to zero.

5.3 Forecaster Information Cost

Rational inattention theory predicts a different measure of relative attention along each dimension, but that each of these measures is parameterized by a single cost of information. With a three

Table 2: Estimates of ω_j, SPF Consensus, Mutual Information

j	10th	20th	30th	40th	50th	60th	70th	80th	90th
ω_j	1.220	1.140	0.829	0.886	0.790	0.580	0.668	0.769	0.701
SE	0.028	0.029	0.027	0.029	0.025	0.025	0.023	0.025	0.023
T-Stat	242.7	173.3	186.3	140.6	116.7	198.0	113.4	134.3	155.1
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The top row shows the estimated cost of information acquisition using 8 estimated moments. These moments correspond to measures of relative attention to the first 8 dimensions of the SPF Consensus forecast. The middle row shows the standard error of the estimated cost, and the bottom row the T-Statistic and associated p-value for the Hansen-Sargan overidentification text, following Galichon and Salanié (2022).

dimensional state space, the attention matrix would be:

$$(\mathbb{I}-\Omega_j) = \begin{pmatrix} 1-f'^{-1}\left(-\frac{\lambda_1}{\omega_j}\right) & 0 & 0 \\ 0 & 1-f'^{-1}\left(-\frac{\lambda_2}{\omega_j}\right) & 0 \\ 0 & 0 & 1-f'^{-1}\left(-\frac{\lambda_3}{\omega_j}\right) \end{pmatrix}$$

Using a minimum distance test, I recover estimates of ω_j assuming different functional forms for $f(\cdot)$. Eigenvalues from the full information benchmark are used for $\lambda_1, \lambda_2, \ldots \lambda_k$. Let $\widehat{1-\Omega_j^i}$ represent the estimated value of the i^{th} value of the attention matrix. The moments to estimate the values of ω_j are estimated using the following moment conditions:

$$\underset{\omega_j}{\text{argmin}} \begin{pmatrix} \left(\widehat{1-\Omega_j^1}\right) - \left(1-f'^{-1}\left(-\frac{\lambda_1}{\omega_j}\right)\right) \\ \left(\widehat{1-\Omega_j^2}\right) - \left(1-f'^{-1}\left(-\frac{\lambda_2}{\omega_j}\right)\right) \\ \cdots \\ \left(\widehat{1-\Omega_j^k}\right) - \left(1-f'^{-1}\left(-\frac{\lambda_3}{\omega_j}\right)\right) \end{pmatrix}$$

Following in Galichon and Salanié (2022), the weight matrix in the generalized method of moment estimation is taken to be the inverse of the variance matrix for the diagonal estimates of $(\mathbb{I}-\Omega_j)$. The implied costs values are summarized in Tables 2 and 3. The size of the first eigenvalue using the SPF as the benchmark is 4.8. The estimated cost of information for the 10th percentile type is about one third of the size of the first eigenvalue, whereas the estimated cost of information for the 90th percentile type is less than one fifth. For the Greenbook sample, the ratios are one-third and one-fourth for the 10th and 90th percentile types, respectively.

Table 3: Estimates of ω_j, Federal Reserve Greenbook, Mutual Information

j	10th	20th	30th	40th	50th	60th	70th	80th	90th
ω_j	1.373	1.137	1.030	0.982	1.004	0.956	0.973	0.962	0.918
SE	0.049	0.048	0.048	0.050	0.045	0.047	0.047	0.044	0.044
T-Stat	31.3	6.7	26.1	11.6	35.4	20.3	18.2	13.0	12.0
p-Value	0.000	0.464	0.000	0.116	0.000	0.005	0.011	0.073	0.102

The top row shows the estimated cost of information acquisition using 8 estimated moments. These moments correspond to measures of relative attention to the first 8 dimensions of the Federal Reserve Greenbook Forecast. The middle row shows the standard error of the estimated cost, and the bottom row the T-Statistic and associated p-value for the Hansen-Sargan overidentification text, following Galichon and Salanié (2022).

5.3.1 Cost of Information with a General Cost Function

I next estimate the parameters for a more general cost function. I assume $f'^{-1}\left(-\frac{\lambda_i}{\omega_j}\right) = \left(\frac{\omega_j}{\lambda_i}\right)^{\alpha_j}$, taking as given the estimated costs ω_j from the previous exercise. Mutual information would correspond to $\alpha_j = 1$. The moment conditions are listed below. Table lists the estimated parameters.

$$\underset{\alpha_j}{\text{argmin}} \begin{pmatrix} \left(\widehat{1-\Omega_j^1}\right) - \left(1 - \left(\frac{\hat{\omega}_j}{\lambda_1}\right)^{\alpha_j}\right) \\ \left(\widehat{1-\Omega_j^2}\right) - \left(1 - \left(\frac{\hat{\omega}_j}{\lambda_2}\right)^{\alpha_j}\right) \\ \cdots \\ \left(\widehat{1-\Omega_j^k}\right) - \left(1 - \left(\frac{\hat{\omega}_j}{\lambda_3}\right)^{\alpha_j}\right) \end{pmatrix}$$

As can be seen from Tables 4 and 5, the assumption of $\alpha = 1$ does not quite fit all types of forecasters. Looking at estimates of the curvature in attention relative to the Federal Reserve Greenbook, there is a general increase in the α_j parameter. Poorly performing forecasters have an estimated α_j below 1, whereas top-performing forecasters have an estimated α_j above 1. Relative to a mutual information benchmark, poorly-performing forecasters display a greater degree of shrinkage on more important factors whereas top-performing forecasters display a lower degree of shrinkage. For higher-order factors this discrepancy with the mutual information benchmark dissipates. However, looking at the estimates of α_j using the SPF Consensus as a benchmark, there is no such pattern. Mutual information appears to be a good fit for low- and top-performing forecasters, whereas forecasted in the middle have relatively more shrinkage on more important factors compared to what would be expected with a mutual information cost function.

Table 6 shows the R^2 values for different estimated models. As can be seen in the table, a the full attention matrix has the best model fit, followed closely by the diagonal attention matrix. Restricting that to instead a mutual information cost function with i^{th} diagonal elements equal to $\left(1 - \frac{\widehat{\omega}_j}{\lambda_i}\right)$ does little to change the fit, suggesting that mutual information is a reasonable model. Estimating the cost of information ω_j using the trace of the precision cost function (i^{th} diagonal elements equal to $\left(1 - \frac{\widehat{\omega}_j}{\lambda_i}\right)^2$) lowers the fit. Indeed, when the cost of information and α parameter

Table 4: Estimates of α_j, SPF Consensus

j	10th	20th	30th	40th	50th	60th	70th	80th	90th
ω_j	1.039	0.925	0.829	0.886	0.790	0.748	0.668	0.769	0.701
SE	0.070	0.074	0.067	0.071	0.060	0.061	0.062	0.069	0.062
α_j	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SE	0.037	0.037	0.032	0.036	0.030	0.030	0.032	0.033	0.031
T-Stat	253.2	174.6	158.7	121.1	103.4	104.7	187.1	142.4	192.8
p-Value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The top row shows the estimated curvature parameter (α_j) of information acquisition function using 8 estimated moments. These moments correspond to measures of relative attention to the first 8 dimensions of the SPF Consensus forecast. The middle row shows the standard error of the estimated cost curvature, and the bottom row the T-Statistic and associated p-value for the Hansen-Sargan overidentification text, following Galichon and Salanié 2022.

Table 5: Estimates of α_j, Federal Reserve Greenbook

j	10th	20th	30th	40th	50th	60th	70th	80th	90th
ω_j	1.153	1.049	0.996	0.933	0.921	0.925	0.943	0.928	0.909
SE	0.099	0.105	0.087	0.093	0.087	0.094	0.087	0.087	0.080
α_j	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SE	0.056	0.060	0.049	0.052	0.048	0.052	0.049	0.048	0.044
T-Stat	127.5	36.3	44.5	19.4	46.7	21.8	27.2	23.2	172.7
p-Value	0.000	0.001	0.000	0.151	0.000	0.083	0.018	0.057	0.000

The top row shows the estimated curvature parameter (α_j) of information acquisition function using 8 estimated moments. These moments correspond to measures of relative attention to the first 8 dimensions of the SPF Consensus forecast. The middle row shows the standard error of the estimated cost curvature, and the bottom row the T-Statistic and associated p-value for the Hansen-Sargan overidentification text, following Galichon and Salanié (2022).

are estimated jointly with the restriction that $\alpha \in [1, 2]$ the boundary of 1 is selected for all forecast ranks.

Table 6: Model Fit, SPF Consensus

Attention Matrix	10th	20th	30th	40th	50th	60th	70th	80th	90th
$H(\mathbb{I} - \Omega_j)H'$	0.297	0.401	0.427	0.419	0.470	0.461	0.526	0.510	0.527
$(\mathbb{I} - \Omega_j)$	0.238	0.354	0.386	0.372	0.441	0.435	0.485	0.488	0.499
Mutual Info	0.213	0.331	0.383	0.366	0.440	0.431	0.485	0.489	0.500
Trace of Precision	0.158	0.310	0.356	0.337	0.422	0.408	0.469	0.474	0.486

The rows display the R^2 for a 20 factor model which assume the following: 1) a model that allows for a non-diagonal attention matrix, 2) estimated diagonal attention matrix, 3) a matrix assuming mutual information ($\alpha = 1$), and 4) a model assuming a trace of the precision ($\alpha = 2$).

Table 7: Model Fit, Federal Reserve Greenbook

Attention Matrix	10th	20th	30th	40th	50th	60th	70th	80th	90th
$H(\mathbb{I}-\Omega_j)H'$	0.251	0.334	0.333	0.386	0.398	0.402	0.412	0.407	0.432
$(\mathbb{I}-\Omega_j)$	0.169	0.254	0.256	0.292	0.322	0.321	0.328	0.331	0.351
Mutual Info	0.153	0.247	0.250	0.291	0.314	0.321	0.330	0.329	0.310
Trace of Precision	0.104	0.230	0.221	0.270	0.296	0.306	0.318	0.316	0.340

The rows display the R^2 for a 15 factor model which assume the following: 1) a model that allows for a non-diagonal attention matrix, 2) estimated diagonal attention matrix, 3) a matrix assuming mutual information ($\alpha = 1$), and 4) a model assuming a trace of the precision ($\alpha = 2$).

5.4 Attention as an Explanation for Forecast Dispersion

The theoretical model forecast model presented in this paper suggests that forecast dispersion is caused by differences in attention to to dimensions of the information space. Attention is reflected both in the relative size of loadings on common factors, as well as the noise in observation error. I look at the ability of the first in explaining forecast dispersion.

Table 8 shows the R^2 from a regression of the observed difference in forecasted values between the 90th and 10th percentile rank forecasts on the model-implied difference in forecasted values for each of the 20 variables at the 6 different horizons using the SPF Consensus as the full information benchmark. For each variable $i = 1, \dots, 20$ and horizon $h = 0, 1, 2, 3, 4, \overline{2Y}$, the R^2 is computed from the regression below, and then further taken as a share of the theoretically explainable variance[15]

$$\tilde{X}^{90}_{i,h,t} - \tilde{X}^{10}_{v,h,t} = \hat{\beta}_{i,h}\hat{F}^*_t \left(\widehat{(\mathbb{I}-\Omega)}_{90} - \widehat{(\mathbb{I}-\Omega)}_{10} \right) \hat{\lambda}^*_{N_i} + u_{i,h,t}$$

The differences in relative loadings explain on average 5.1% of observed dispersion, and at most 19% (three-quarter ahead GDP Price Index). The amount of explained dispersion is in general increasing with forecast horizon (3.1% for this quarter and 6.4% for next quarter). Attention can explain between 17.1 and 19.2% of forecast dispersion of the GDP Price Index of forecast horizons between 1 and 4 quarters. For CPI, attention explains about 10% of the dispersion in one and two quarter ahead forecasts. Similarly for real GDP, attention can explain 17.5% of the dispersion in forecasts for four quarters ahead. A diagonal measure of attention, largely parameterized by a single sensitivity parameter explains a sizeable portion of forecast dispersion.

Table 9 shows results using the Federal Reserve Greenbook as the full information benchmark. The results are similar. Attention on average explains more of the dispersion for forecasts than for Nowcasts, with the peak occurring at the two-quarter ahead forecast.

[15] This excludes the variance in the idiosyncratic error of signals.

Table 8: Forecast Dispersion Explained, SPF Consensus, 90th - 10th Percentile

Variable	t+0	t+1	t+2	t+3	t+4	t+2Y	Mean
RGDP	0.018	0.069	0.063	0.063	0.175	0.076	0.077
NGDP	0.050	0.102	0.052	0.069	0.079	0.095	0.075
CPROF	-0.000	0.025	0.011	0.043	0.016	0.010	0.017
HOUSING	0.034	0.035	0.042	0.051	-0.000	0.008	0.028
INDPROD	0.022	0.159	0.024	0.042	0.107	0.105	0.077
PGDP	0.037	0.171	0.183	0.192	0.184	0.074	0.140
RCONSUM	0.026	0.155	0.091	0.167	0.057	0.088	0.097
RFEDGOV	0.002	0.008	0.000	0.001	0.004	0.044	0.010
RNRESIN	0.102	0.032	0.061	0.005	0.104	0.038	0.057
RRESINV	0.010	0.083	0.011	0.017	0.003	0.007	0.022
RSLGOV	0.054	0.048	0.015	0.001	0.065	0.065	0.041
UNEMP	0.054	0.038	0.055	0.051	0.040	0.072	0.051
EMP	0.033	0.087	0.052	0.001	0.005	-0.004	0.029
RCBI	0.078	0.046	0.046	0.039	0.078	0.036	0.054
REXPORT	0.027	0.013	0.070	0.018	0.068	0.025	0.037
TBILL	0.014	0.014	0.088	0.131	0.096	0.125	0.078
TBOND	-0.003	0.002	0.002	0.081	0.014	0.044	0.023
BOND	0.060	0.064	0.065	0.056	0.038	-0.019	0.044
TSPREAD	-0.001	0.010	0.024	0.019	0.007	-0.000	0.010
CPI	0.000	0.119	0.092	0.057	0.052	0.051	0.062
Mean	0.031	0.064	0.052	0.055	0.060	0.047	0.051

The rows display the R^2 from regressing the observed difference between forecasted variables by the 90th and 10th percentile ranked forecasters on the model implied difference in forecasts. Forecast models include 20 factors.
For each variable $i = 1, \ldots, 20$ and horizon $h = 0, 1, 2, 3, 4, \overline{2Y}$, the R^2 is computed from the following regression, scaled by the total possible variance explained:

$$\tilde{X}^{90}_{i,h,t} - \tilde{X}^{10}_{v,h,t} = \hat{\beta}_{i,h} \hat{F}^*_t \left((\widehat{\mathbb{I} - \Omega})_{90} - (\widehat{\mathbb{I} - \Omega})_{10} \right) \hat{\lambda}^*_{N_i} + u_{i,h,t}$$

Table 9: Forecast Dispersion Explained, Federal Reserve Greenbook, 90th - 10th Percentile

Variable	t+0	t+1	t+2	t+3	t+4	Mean
RGDP	0.026	0.087	0.127	0.071	0.147	0.092
NGDP	0.046	0.064	0.091	0.124	0.171	0.099
HOUSING	0.000	0.021	0.007	0.092	0.004	0.025
INDPROD	0.025	0.036	0.040	0.022	0.028	0.030
PGDP	0.071	0.120	0.024	0.036	0.112	0.073
RCONSUM	0.009	0.129	0.087	0.127	0.078	0.086
RFEDGOV	0.011	0.039	0.007	0.025	0.001	0.016
RNRESIN	0.046	0.035	0.031	0.005	0.004	0.024
RRESINV	0.025	0.103	0.018	0.063	0.001	0.042
RSLGOV	0.009	0.021	0.041	0.006	0.030	0.021
UNEMP	0.001	0.009	0.003	0.009	0.031	0.011
TBILL	0.006	0.000	0.016	0.001	0.002	0.005
TBOND	0.007	0.027	0.019	0.013	0.035	0.020
CPI	0.000	0.035	0.035	0.032	0.028	0.026
Mean	0.020	0.052	0.039	0.045	0.048	0.041

The rows display the R^2 from regressing the observed difference between forecasted variables by the 90th and 10th percentile ranked forecasters on the model implied difference in forecasts.
Forecast models include 15 factors.
For each variable $i = 1, \ldots, 14$ and horizon $h = 0, 1, 2, 3, 4$, the R^2 is computed from the following regression, scaled by the total possible variance explained:

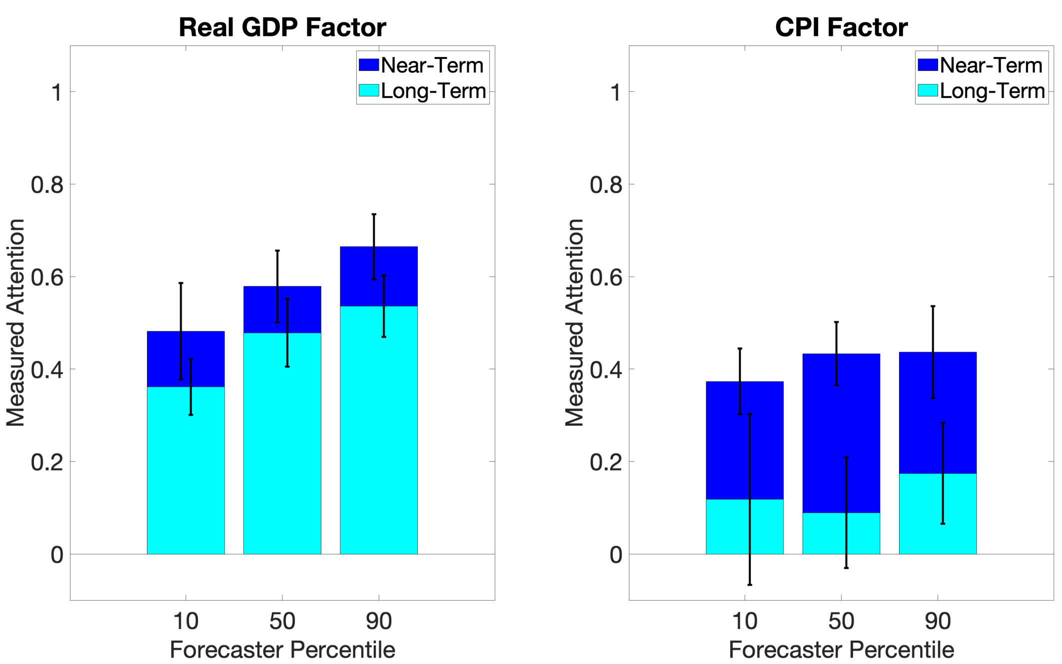
$$\tilde{X}^{90}_{i,h,t} - \tilde{X}^{10}_{v,h,t} = \hat{\beta}_{i,h} \hat{F}^*_t \left((\widehat{\mathbb{I} - \Omega})_{90} - (\widehat{\mathbb{I} - \Omega})_{10} \right) \hat{\lambda}^*_{N_i} + u_{i,h,t}$$

5.5 Attention to Economically-Meaningful Factors

If the forecaster's economic model satisfies the assumption of Lemma 3 (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.

Suppose that the economic forecasts are driven by 4 latent factors that 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. This ordering would be consistent with the logic from the start of section 4. Measured attention along the rotated factors is captured in Figure 12.

Figure 12: Measured Attention to Near- and Long-Term Factors



Measured attention by forecaster type to orthogonal factors ordered: near-term real GDP growth, near-term CPI growth, long-term GDP growth, long-term CPI growth. Factors capture comovements orthogonal to previous factors.

Under the assumption of the underlying forecast model, forecasters at the 10th and 50th percentile are shown to pay no significant attention to the long-term CPI factor. Forecasts of 10th and 50th percentile forecasters underreact to changes in the the factor associated with changes specific to the SPF Consensus long-term CPI forecasts.

6 Discussion and Conclusion

There are many theoretical reasons that forecasts from different agents might not be the same - the agents might have different forecast models or might display persistent optimistic or pessimistic biases. In this paper I explore differences in forecasts which reflect the intensity of forecaster attention.

In the sets of forecasts I consider, time series panels associated with forecast performance type, forecasters share a similar reduced form model. The largest comovements in all of the forecast panels are explained by factors associated with similar groups of variables. However, the degree to which these common factors are mapped into forecasted variables is relatively smaller when compared with proxies for full information forecasts. Professional macroeconomic forecasts exhibit shrinkage, which, through the lens of a rational inattention model, can be interpreted as relative attention.

The multidimensional measures of relative attention are consistent with predictions from a multivariate rational inattention model. Attention is in general lower for less important factors. Minimum distance tests on the multidimensional measure of relative attention show that a model with a single cost parameter (ω_j) can explain a large portion of forecast variance. These differences in attention can in turn explain a modest proportion of forecast dispersion.

It is important for policymakers to take into account the multivariate nature of forecasts and what that implies about optimal attention allocation in models of expectation formation.

References

H. Afrouzi and C. Yang. Dynamic rational inattention and the phillips curve. *Working Paper.*, 2021.

P. Andrade, R. K. Crump, S. Eusepi, and E. Moench. Fundamental disagreement. *Journal of Monetary Economics*, 83:106–128, 2016. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2016.08.007. URL https://www.sciencedirect.com/science/article/pii/S0304393216300794.

J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/3082043.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/2692167.

J. Bai and S. Ng. Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics*, 131(1):507–537, 2006. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2005.01.015. URL https://www.sciencedirect.com/science/article/pii/S0304407605000205.

J. Bai and S. Ng. Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1):18–29, 2013. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2013.03.007. URL https://www.sciencedirect.com/science/article/pii/S0304407613000651.

J. Bai and S. Ng. Rank regularized estimation of approximate factor models. *Journal of Econometrics*, 212(1):78–96, 2019. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2019.04.021. URL https://www.sciencedirect.com/science/article/pii/S0304407619300764. Big Data in Dynamic Predictive Econometric Modeling.

F. Bianchi, S. C. Ludvigson, and S. Ma. Belief distortions and macroeconomic fluctuations. *American Economic Review*, 112(7):2269–2315, July 2022. doi: 10.1257/aer.20201713. URL https://www.aeaweb.org/articles?id=10.1257/aer.20201713.

E. Cahan, J. Bai, and S. Ng. Factor-based imputation of missing values and covariances in panel data of large dimensions. *Journal of Econometrics*, 233(1):113–131, 2023. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2022.01.006. URL https://www.sciencedirect.com/science/article/pii/S0304407622000215.

J.-F. Cai, J. Candès, Emmanuel, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

A. Caplin, M. Dean, and J. Leahy. Rationally inattentive behavior: Characterizing and generalizing shannon entropy. *Journal of Political Economy*, 130(6):1676–1715, 2022.

O. Coibion and Y. Gorodnichenko. Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8):2644–78, August 2015. doi: 10.1257/aer.20110306. URL https://www.aeaweb.org/articles?id=10.1257/aer.20110306.

T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons, second edition, 2006.

M. Dean and N. N. Neligh. Experimental tests of rational inattention. *Journal of Political Economy*, Forthcoming. doi: 10.1086/725174. URL https://doi.org/10.1086/725174.

J. Dovern. A multivariate analysis of forecast disagreement: Confronting models of disagreement with survey data. *European Economic Review*, 80:16–35, 2015. ISSN 0014-2921. doi: https://doi.org/10.1016/j.euroecorev.2015.08.009. URL https://www.sciencedirect.com/science/article/pii/S0014292115001257.

C. Fulton. Mechanics of linear quadratic gaussian rational inattention tracking problems. *Finance and Economics Discussion Series*, 2017.

A. Galichon and B. Salanié. Estimating separable matching models. *Working Paper*, 2022.

M. Giacoletti, K. T. Laursen, and K. J. Singleton. Learning from disagreement in the u.s. treasury bond market. *The Journal of Finance*, 76(1):395–441, 2021. doi: https://doi.org/10.1111/jofi.12971. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12971.

D. Giannone, L. Reichlin, and D. Small. Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676, 2008. ISSN 0304-3932. doi: https://doi.org/10.1016/j.jmoneco.2008.05.010. URL https://www.sciencedirect.com/science/article/pii/S0304393208000652.

B. Hébert and M. Woodford. Neighborhood-based information costs. *American Economic Review*, 111(10):3225–55, October 2021. doi: 10.1257/aer.20200154. URL https://www.aeaweb.org/articles?id=10.1257/aer.20200154.

E. P. Herbst and F. Winkler. The factor structure of disagreement. *Board of Governors of the Federal Reserve System (U.S.)*, 2021.

M. Kacperczyk, S. V. Nieuwerburgh, and L. Veldkamp. A rational theory of mutual funds' attention allocation. *Econometrica*, 84(2):571–626, 2016. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/43866444.

B. Kőszegi and F. Matějka. Choice Simplification: A Theory of Mental Budgeting and Naive Diversification*. *The Quarterly Journal of Economics*, 135(2):1153–1207, 01 2020. ISSN 0033-5533. doi: 10.1093/qje/qjz043. URL https://doi.org/10.1093/qje/qjz043.

N. G. Mankiw, R. Reis, and J. Wolfers. Disagreement about inflation expectations. *NBER Macroeconomics Annual*, 18:209–248, 2003. doi: 10.1086/ma.18.3585256. URL https://doi.org/10.1086/ma.18.3585256.

C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50(3):665–690, 2003. ISSN 0304-3932. doi: https://doi.org/10.1016/S0304-3932(03)00029-1. URL https://www.sciencedirect.com/science/article/pii/S0304393203000291. Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179, 2002a. ISSN 01621459. URL http://www.jstor.org/stable/3085839.

J. H. Stock and M. W. Watson. Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162, 2002b. ISSN 07350015. URL http://www.jstor.org/stable/1392055.

Appendix

Proof of Lemma 2

This proof follows that in Afrouzi and Yang (2021). Let u_i^B characterize the eigenvector corresponding to the i^{th} largest eigenvalue of the BB' matrix (the i^{th} column of U^B). The optimal choice of posterior variance can be written as:

$$\begin{aligned}\mathbb{I} - \bar{P}^* &= U^B \left(\mathbb{I} - \Omega\right) U^{B\prime} \\ &= \sum_{i}^{n} \max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right) u_i^B u_i^{B\prime}\end{aligned}$$

The agent has Gaussian prior beliefs about the variance of fundamental shocks and can only acquire Gaussian signals. With a choice of Z and H, the agent's posterior belief posterior belief, $\bar{P}$, is given by:

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

Rearranging and replacing the posterior variance with the choice of optimal yields:

$$\sum_{i=1}^{n} \max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right) u_i^B u_i^{B\prime} = Z' \left(ZZ' + H\right)^{-1} Z$$

Choose a signal structure such that $Z'Z = 1$, this additionally implies a diagonal H matrix. Let h_i indicate the i^{th} diagonal element of H, and z_i indicate the i^{th} column of z. The equality can then be rewritten as:

$$\sum_{i=1}^{n} \max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right) u_i^B u_i^{B\prime} = \sum_{i=1}^{n} \left(1 + h_i\right)^{-1} z_i' z_i$$

One way for this inequality to hold is if $z_i' = u_i^B$ for $i \in 1, \ldots n$. Additionally, for a given i, if $\max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right) = 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)$, then $h_i = \left(1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right)^{-1} - 1$. If $\max\left(0, 1 - f'^{-1}\left(-\frac{\lambda_i^B}{\omega}\right)\right) = 0$ then the agent does not acquire a signal along that dimension (intuitively: signal error variance $\to \infty$).

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.

$$\Phi = H\left(\mathbb{I} - \Omega\right)H^{-1}$$

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:

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

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.

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

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

Second stage coefficients:

Take the k^{th}, ($k \in 1,\ldots r$) coefficient for each of the N forecasted variables (denoted by $\widehat{(\varPhi\lambda_i^*)}_k$), and regress it on the full set of loadings.

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

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

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

By Assumption 6, the middle term converges to $\Sigma_{\Lambda e F_k}$. 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 $\varPhi$ is:

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

Description of Monte Carlo Simulations

The state of the economy evolves according to the following latent factor model:

$$f_t = Af_{t-1} + z_t$$

Agent j makes a forecast of a vector of forecasts at time t based on their belief about the realization of the shock variable z_t, $\hat{z}_{jt}$, as well as the common belief about $\hat{f}_{t-1}$[16].

$$\hat{x}_{jt} = B\left(A\hat{f}_{t-1} + \hat{z}_{jt}\right)$$

Consider signals z_t. Agents can choose the variance of those signals in order to maximize their utility. Following Afrouzi and Yang, these optimal set of signals chosen by the agents will be spanned by dimensions of

$$s_{jit} = U_i^{B\prime} z_{jit} + \eta_{jit} \qquad \eta_{jit} \sim N\left(0, \sigma_{ji\eta}\right)$$
$$\sigma_{ji\eta} = \frac{\omega_j}{D_i - \omega_j}$$

Start with the case in which z_t is iid with 0 mean and unit variance. Following receipt of the vector of signals, $s_{jt} = (s_{j1t}, \ldots, s_{jNt})'$, agents update their belief of the i^{th} dimension of $\hat{z}_{jt}$ to be:

$$\hat{z}_{jit} = U_i^B\left(1 + \frac{\omega_j}{D_i - \omega_j}\right)^{-1} U_i^{B\prime}\left(z_{jit} + \eta_{jit}\right)$$

In matrix form:

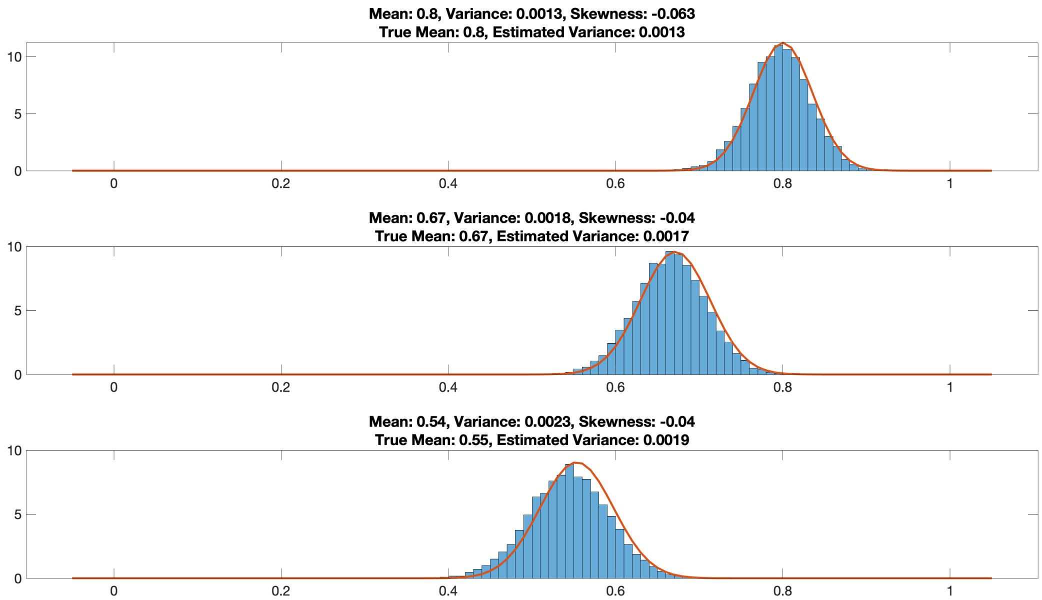
$$\hat{z}_{jt} = U^B\left(\mathbb{I} - \Omega_j\right)^{-1} U^{B\prime}\left(z_{jt} + \eta_{jt}\right)$$

The agent's forecast is then:

$$\hat{x}_{jt} = BA\hat{f}_{t-1} + B\left(\mathbb{I} - \Sigma_j\right)^{-1}\left(z_{jt} + \eta_{jt}\right)$$

[16]Releases of major aggregate data (i.e. GDP) provide the best signal of these latent factors.

Low Cost



High Cost