Macroeconomic Expectations and Cognitive Noise[*]

Yeji Sung[†]

Columbia University

Click for the latest version

December 17, 2022

Abstract

Conventional models of information frictions assume it is costly to process external information. However, these models cannot explain puzzling patterns observed in survey forecasts. I propose a model in which *internal* information — knowledge stored in memory — is also costly to process. The model is consistent with survey-forecast patterns and offers an estimation strategy to identify the extent of information frictions. I then explore the macroeconomic implications of these frictions. The proposed model suggests that inflation expectations are not well anchored, making it more challenging to stabilize inflation than under conventional information-friction models.

[*]I am grateful to Michael Woodford, Hassan Afrouzi, Jennifer La'O, and Mark Dean for their guidance and support. This paper has benefited from discussions with members of the Cognition and Decision Lab and seminar participants at Columbia University, the Federal Reserve Board, and the Federal Reserve Bank of St. Louis. I owe special thanks to Bill Dupor, Benjamin Johannsen, Fabian Winkler, Giovanni Nicolò, Fernando Martin, Yueran Ma, Rava Azeredo da Silveira, and Miguel Acosta for their insightful feedback.

[†]Department of Economics, Columbia University. Email:yeji.sung@columbia.edu.

1 Introduction

Despite being commonly held, the empirical validity of the full-information rational expectations (FIRE) assumption has long been questioned. According to FIRE, forecasts made by economic agents should not include predictable errors because all information is used efficiently. However, an expanding empirical literature has found that this prediction does not hold, even for professional forecasters who presumably have ample information and advanced tools.

In particular, recent studies argue that econometricians can predict errors that forecasters will make based on the latter's recent forecast revisions. However, a puzzle emerges. Coibion and Gorodnichenko (2015) finds that the average forecast tends to *undershoot* realizations when forecasters revise their projections upward. In comparison, Bordalo, Gennaioli, Ma, and Shleifer (2020b) find that when an individual forecaster revises her projection upward, her forecast tends to *overshoot* realizations. Of course, the revision of the average forecast is more muted than the revision of the individual one, as the average forecast averages different views. However, it is not apparent how to explain the flip in the direction of predictability.

This paper shows that noisy information can account for the seemingly contradictory pattern. While forecasters have access to a vast amount of information, they have finite capacity to process it. I propose that forecasts are based on the *mental representation* of available information, not all available information. The mental representation can be considered a noisy summary of information; it is composed of both the mind's distilled understanding of available information and the noise uncorrelated with the distillation. This process is endogenously determined to maximize forecast accuracy given processing constraints.

To explain these patterns in both average and individual forecasts, I distinguish between two types of information: external and internal. External information can be looked up; it includes data releases, news articles, press conferences, and the like. Internal information is stored in forecasters' memory; it consists in their accumulated knowledge from past forecasting experience. Importantly, both types of information are noisy: forecasts are formed based on the mental representation of available external and internal information. Using this model, I estimate the extent of information frictions using professional forecasters' projections of the overall US economy. Furthermore, I explore the monetary-policy implications of the estimated model.

Conventional models of information frictions assume that only external information is noisy. These models can explain why consensus forecasts undershoot. When facing a new set of external information, forecasters, on average, are less responsive than under FIRE because the new information is noisy. At the individual level, forecasters make projections efficiently given the noise. Therefore, such revisions do not predict systematic errors in individual forecasts.

I show that by adding noisy internal information, I can explain the predictability of both average and individual forecasts. Noisy external information generates the consensus-level pattern, as in Coibion and Gorodnichenko (2015), while noisy internal information generates the individual-level pattern. When forecasters cannot freely access their internal information, their prior knowledge resolves less uncertainty about the forecast variable. Thus, forecasters put extra weight on new information. This extra sensitivity to new information explains why individual forecasts tend to overshoot more than conventional information-friction models predict.

Jointly considering both types of noisy information is crucial for understanding the extent of information frictions. This is because the sensitivity with which forecasts are revised depends on both types of information. Forecast revisions can be sensitive for two reasons: external information is not very noisy, or internal information is quite noisy. I show that the extent of the two types of noise determines the revision pattern of average and individual forecasts. Not considering both types of noisy information would lead one to misestimate the degree of information frictions.

A direct implication is that Coibion and Gorodnichenko (2015) underestimates the extent of information frictions. The authors argue that the severity of information constraints can be inferred from the revision pattern of consensus forecasts. However, since this methodology implicitly assumes that accessing internal information is costless, it does not account for the extra weight on new information arising from noisy internal information. Therefore, the methodology proposed by Coibion and Gorodnichenko (2015) misattributes this extra sensitivity to less severe information frictions.

To improve the model's empirical validity, I extend the model along one more dimension: forecasters learn about the long-run mean. Instead of assuming that forecasters are fully aware of where the forecast variable reverts to, I assume they learn about the long-run steady state over time. While more than one parameter determines the steady state, I focus on the mean because knowledge about the mean is essential for forecasters making long-term forecasts.

In the proposed model, forecasters are perpetually uncertain about the long-run mean. If accessing internal information is costless, forecasters eventually learn about the mean. This is the basis on which many models assume forecasters have perfect awareness of the model parameters. However, as my other work Azeredo da Silveira et al. (2020) shows, forecasters' knowledge about the mean is imperfectly accumulated over time when internal information is not perfectly accessible. In this case, learning persists even after extensive learning opportunities.

Using the extended model, I estimate the degree of information frictions. The model is applied to professional forecasters' projections of US economic variables related to output, the price level, the labor and housing markets, and borrowing costs. For each macroeconomic variable, the two constraints — one in processing external information and the other in processing internal information — are inferred from the forecast-revision patterns at the consensus and individual levels. I find that the extent of information frictions is more substantial than what the conventional information-frictions literature finds: my estimate of the constraint in processing external information is twice as large as that of Coibion and Gorodnichenko (2015). I also show that the estimated model explains sizable shares of the variation in forecasts and revisions, both in the cross sections and in the time series.

Then, I explore the macroeconomic implications of the proposed information frictions — in particular, the implications for how inflation is determined. I use a standard New Keynesian model in which firms set prices based on their macroeconomic expectations. Using this framework, I show how the inflation process varies with the assumption of the expectation-formation process. Furthermore, I investigate the operation of monetary policy in balancing the trade-off between inflation and output stabilization.

If firms are subject to the costly information proposed in this paper, stabilizing inflation can be more challenging than under FIRE. The key reason is that inflation expectations are unanchored because internal information is costly to process. Since firms do not have perfect awareness of the long-run economy, their beliefs about it fluctuate with persistence. This additional fluctuation is transmitted through their price setting, making aggregate inflation volatile. In this economy, a monetary policy strongly emphasizing inflation stabilization can more effectively guide economic agents' long-run expectations.

In proposing a new expectation-formation model, I provide a parsimonious explanation for the puzzling features of survey forecasts. In the model I present, one type of information friction keeps economic agents from making forecasts consistent with FIRE: finite capacity to process a vast amount of available information. In comparison, previous proposals in the literature resort to a non-Bayesian assumption *in addition to* information frictions to explain the forecast-revision patterns discussed in this paper. For example, representative heuristics (Bordalo, Gennaioli, Ma, and Shleifer (2020b)), misspecification of the model (Angeletos, Huo, and Sastry (2021)), and desire to stand out from the crowd (Gemmi and Valchev (2021)) have been proposed. While these may be plausible and insightful proposals, it is unclear how economic agents come to have such biases. Furthermore, I show that the model I present explains features of survey forecasts regarding the forecast horizon that these previous proposals cannot explain.

The findings from this paper also shed light on the formation of *long-run* inflation expectations. It has long been recognized that economic agents' inflation expectations affect the inflation process. Thus, the implementation of monetary policy should carefully consider the exact nature of expectation formation (Orphanides and Williams (2004)). However, recent literature argues that expectations about long-run inflation are crucial to understanding past inflation dynamics (Carvalho, Eusepi, Moench, and Preston (2022), Hazell, Herreño, Nakamura, and Steinsson (2022)) and have important monetary-policy implications (Gàti (2021)). A popular proposal in the literature is that economic agents learn about unobservable stochastic trends from the current economy (K. Crump, Eusepi, Moench, and Preston (2021), Farmer, Nakamura, and Steinsson (2021)). While plausible, this idea predicts that economic agents should have well-anchored long-run expectations in response to a sudden spike in inflation if they have experienced low and stable inflation for a long period. In this paper, seemingly anchored long-run inflation expectations can start moving when agents witness bouts of high inflation. This prediction is consistent with experimental studies documenting fluctuations in long-term beliefs in a stable-trend environment (Afrouzi, Kwon, Landier, Ma, and Thesmar (2020)).

More generally, this paper contributes to our understanding of how cognitive limitations affect economic agents' beliefs and decisions. In various fields of economics, cognitive limitations have been proposed to explain several seemingly unrelated patterns (Woodford (2020)). In macroeconomics, rational-inattention theories have been proposed to explain why macroeconomic variables respond to fluctuations in the economy (for example, monetary-policy shocks) with a long delay (Sims (2003), Mackowiak and Wiederholt (2009)). In behavioral economics, cognitive uncertainty has been proposed as a unifying explanation for several patterns often viewed as distinct phenomena (Enke and Graeber (2019)). I contribute to this literature by showing that cognitive limitations help us understand the puzzling patterns of survey forecasts emphasized in the macroeconomic literature on expectations.

The paper proceeds as follows. Section 2 presents a model of expectation formation in which forecasts are based on the mental representation of available information. Section 3 discusses what representation is optimal given the information constraints. Section 4 presents the model prediction of the forecast-revision patterns and the estimation strategy. Section 5 describes an extension of the expectation model. Section 6 presents the structural-estimation results. Section 7 describes the illustrative macroeconomic model and discusses the monetary-policy implications. Section 8 concludes.

2 A Model of Mental Representation

In this section, I introduce a model of mental representation (that is, a noisy summary of available information). I describe how a vast amount of information is processed and stored in memory.

2.1 The Forecasting Problem

Consider macroeconomic variable y_t, which is the sum of persistent and transitory components. I assume that

$$y_t = z_t + \eta_t,$$

where η_t is i.i.d, and

$$z_t = (1-\rho)\,\mu + \rho\, z_{t-1} + \epsilon_t,$$

where μ is the long-run mean of z_t, ρ is the serial correlation of z_t (with $|\rho| < 1$), and ϵ_t is an i.i.d. sequence drawn from Gaussian distribution $\mathcal{N}\left(0,\, \sigma_\epsilon^2\right)$. I assume y_t is observable but z_t and η_t are not directly observed. I assume that all values of parameters describing the stochastic process are known.

The forecasters' problem is to produce projections for future realizations of y_t. The loss from incorrectly forecasting is described by the expected value of a quadratic loss function:

$$E\left[\sum_{t=0}^{\infty} \beta^t \sum_{h=1}^{H} \left(y_{t+h} - F_{i,t}\, y_{t+h}\right)^2\right] \tag{1}$$

Here, $F_{i,t}\, y_{t+h}$ is decision-maker (DM) i's forecast of y_{t+h}. Forecasters make projections up to H

periods ahead. The expectation operator E is over every possible piece of information available at time t and is described in the remaining section.

Available information. I categorize information into two types: external and internal. Forecasters can look up external information. It includes quantitative and qualitative information, such as data releases, press conferences, and market reports. Internal information is in forecasters' memory —- that is, their past cognitive state.

Cognitive constraints: mental representation. Forecasters' external and internal information is high dimensional and complex. They have a finite capacity to process such data. To capture this constraint, I introduce the notion of mental representation from psychology and cognitive science; it can be considered a noisy summary of information. I propose that forecasters base their projections on the mental representation of available information instead of all available information. The original complex data is distilled into a simpler form and compounded with random noise, which makes the representation imprecise. This representation is optimally determined, as discussed below.

2.2 Mental Representation of External Information

External information. The underlying state z_t is partially revealed by many pieces of quantitative and qualitative information. Examples of quantitative information are historical realizations of past y_t or other variables relevant for predicting z_t. Qualitative information includes opinions and market commentaries. All such information that is at least somewhat informative about the value of z_t is stored in a large vector N_t. The relationship between N_t and z_t is described as follows:

$$N_t = R \cdot z_t + \nu_t \tag{2}$$

R is a constant vector, and $\nu_t \sim \mathcal{N}(0,\, V)$ for some positive definite matrix V.

Imprecise representation. DM uses various kinds of information in N_t when making forecasts of y_t. I assume that how precisely DM's forecasts depend on this external information is constrained. In particular, I assume that knowledge from N_t is described as follows:

$$n_{i,t} = K_t \cdot N_t + u_{i,t} \tag{3}$$

Here, K_t is a matrix (possibly with many fewer rows than the number of elements in N_t) and $u_{i,t} \sim \mathcal{N}(O,\, \Sigma_{u,t})$ for some positive semidefinite matrix $\Sigma_{u,t}$. The noise $u_{i,t}$ is not correlated with z_t and is idiosyncratic to each forecaster.

The matrices K_t and $\Sigma_{u,t}$ are endogenously determined subject to a constraint. The degree of precision of the mental representation $n_{i,t}$ is measured with the Shannon mutual information between $n_{i,t}$ and N_t, denoted as $\mathcal{I}(n_{i,t};\, y_t)$.[1] More inaccurate representation is captured by lower mutual information between the two random variables. I assume that the precision of mental representation is constrained as follows:[2]

$$I(n_{i,t}; N_t) \leq -\frac{1}{2}\ln\phi_n \tag{4}$$

Here, $\phi_n \in (0,1)$ parameterizes the upper bound of the mutual information that is taken as given. One can see that a higher ϕ_n allows lower mutual information, thereby constraining the accuracy of the mental representation.

If $\phi_n \to 0$, then forecasts are accurately based on information in N_t. In this case, K_t is an identity matrix (whose dimension is equivalent to the number of rows in N_t) and $\Sigma_{u,t}$ is a zero matrix (with the same dimension as K_t). With $\phi_n > 0$, forecasts are based on the approximate representation of N_t, as K_t may have many fewer rows than the number of elements in N_t and at least some of the diagonal elements of $\Sigma_{u,t}$ are positive. When $\phi_n \to 1$, forecasts are not based on information in N_t, since the representation is infinitely inaccurate.

2.3 Mental Representation of Internal Information

Internal information. In addition to external information N_t, I assume that DM has access to internal information such as her past cognitive state. I denote the internal information accessible at t as $(m_{i,t-1}, n_{i,t-1})$. As discussed earlier, $n_{i,t-1}$ is the mental representation of the news vector N_{t-1}. Meanwhile, $m_{i,t-1}$ is the knowledge carried through $t-1$ before observing N_{t-1}. One can think of $m_{i,t-1}$ as the memory stock of knowledge, and its evolution will be discussed shortly.

Imperfect representation. I assume the internal information can be represented as follows:

$$m_{i,t} = \Lambda_t \cdot \begin{pmatrix} m_{i,t-1} \\ n_{i,t-1} \end{pmatrix} + \omega_{i,t} \tag{5}$$

Here, Λ_t is a matrix that may have fewer rows than $(m_{i,t-1}, n_{i,t-1})$ and $\omega_{i,t}$ is an i.i.d. sequence that is uncorrelated with $(m_{i,t-1}, n_{i,t-1})$ and drawn from the Gaussian distribution $\mathcal{N}(O,\, \Sigma_{\omega,t})$ for some positive semidefinite matrix $\Sigma_{\omega,t}$.

The two matrixes Λ_t and $\Sigma_{\omega,t}$ are chosen optimally subject to the constraint. The extent of noise in the mental representation $m_{i,t}$ is measured with the Shannon mutual information between

[1] This metric captures how "close" $n_{i,t}$ is to N_t. If $\mathcal{I}(n_{i,t};\, y_t)$ is close to zero, then it means knowing $n_{i,t}$ is not informative about N_t. If, on the other hand, the metric is close to infinity, then information delivered by $n_{i,t}$ about N_t is perfectly accurate.

[2] The proposed cost function is different from what is typically assumed in the rational-inattention literature. There, it is assumed that DM can arrange to receive a signal $n_{i,t}$ at time t, conditioning on all the signals till time $t-1$. That is, the cost is assumed to be proportional to $\mathcal{I}(n_{i,t}; y_t | n_{i,t-1}, \cdots, n_{i,0})$. As will be clear from the rest of the model, I consider an environment in which the past realized values of $n_{i,t}$ are not freely available. Therefore, I assume that external information is processed independently of the cognitive state.

$m_{i,t}$ and $(m_{i,t-1}, n_{i,t-1})$. The lower mutual information captures a more inaccurate representation of internal information. In parallel with (4), I assume that the accuracy of the representation is constrained as follows:

$$I\left(m_{i,t}; m_{i,t-1}, n_{i,t-1}\right) \leq -\frac{1}{2} \ln \phi_m \tag{6}$$

Here, $\phi_m \in (0,1)$ is taken as given. A higher ϕ_m means a more constrained representation.

If $\phi_m \to 0$, forecasts are accurately based on internal information. The corresponding mental representation is when Λ_t is an identity matrix and $\Sigma_{\omega,t}$ is a zero matrix. With $\phi_m > 0$, forecasts rely on imperfect representation of $(m_{i,t-1}, n_{i,t-1})$. When $\phi_m \to 1$, forecasts are not based on internal information, since the represented information is completely inaccurate.

2.4 Forecasts Based on Represented Information

We have seen how external and internal information is mentally represented. For brevity, I refer to $n_{i,t}$ as *noisy news* (that is, an imperfect representation of external information) and $m_{i,t}$ as *noisy memory* (that is, an imperfect representation of internal information). I consider the representation to be noisier if the accuracy of representation is more constrained (higher ϕ_n or ϕ_m).

Bayesian forecasts subject to information constraints. I assume that forecasts are Bayesian efficient given the noisy news $n_{i,t}$ and noisy memory $m_{i,t}$. That is, while the bottleneck is in processing the complex, high-order information, forecasters have expertise in combining $n_{i,t}$ and $m_{i,t}$. The conditional distribution is derived using the usual Kalman filter formula.

Implications of the linear-Gaussian structure. The linear-Gaussian structure of $n_{i,t}$ and $m_{i,t}$ implies that DM's beliefs about the past and current realizations of z_t take the form of a Gaussian distribution. In other words, $(z_0, \cdots, z_t) | m_{i,t}$ and $(z_0, \cdots, z_t) | m_{i,t}, n_{i,t}$ are both Gaussian. (The second moment of the Gaussian distribution captures the uncertainty DM feels, which depends on the severity of news noise and memory noise.) Since DM's beliefs about the past and current realizations are Gaussian, DM's belief about future realizations is also Gaussian.

I introduce the following notations to denote DM's beliefs about the state z_τ implied by her cognitive states:

$$z_\tau | \, m_{i,t} \sim \mathcal{N}\left(z^m_{i,\tau|t}, \, \Sigma^m_{\tau|t}\right)$$
$$z_\tau | \, m_{i,t}, n_{i,t} \sim \mathcal{N}\left(z_{i,\tau|t}, \, \Sigma_{\tau|t}\right)$$

The top distribution refers to the (beginning of period t) prior belief conditioned on the memory state at time t. The superscript m indicates that beliefs are based on memory alone. The bottom distribution is the posterior belief after observing $n_{i,t}$ (and is denoted without the superscript m).

Then, the optimal forecasts of y_{t+h} will be

$$F_{i,t}\, y_{t+h} = \left(1 - \rho^h\right)\mu + \rho^h\, z_{i,t|t},$$

from which the mean squared error from forecasting y_{t+h} equals

$$E\left[\left(y_{t+h} - F_{i,t}\, y_{t+h}\right)^2\right] = \rho^{2h}\, \Sigma_{t|t},$$

where the expectation is over the entire joint probability distribution of possible values of z_t, $m_{i,t}$, and $n_{i,t}$. The average losses from inaccurate forecasting are proportional to $\Sigma_{t|t}$. The loss function (1) then reduces to

$$\sum_{t=0}^{\infty} \beta^t \left[q \cdot \Sigma_{t|t}\right], \tag{7}$$

where $q \equiv \frac{\rho^2(1-\rho 2H)}{1-\rho^2}$ is a constant known to DM.

2.5 The Nature of Information Frictions

In conventional models of information frictions, forecasters have noisy (or dispersed) information about the state of the economy because they observe the state with idiosyncratic errors (Woodford (2003)). The usual interpretation of this assumption is that forecasters have some fragmented information about the state and no one knows the state perfectly. It is typical to assume further that forecasters store their information and access it in any future period.

In contrast, the information friction in this paper is a cognitive constraint. It is not that forecasters have different sources of information per se but that random cognitive noise enters while processing the vast set of information. Therefore, even with access to the same information, forecasters have a somewhat different understanding or interpretation of the data, as in Sims (2003).

Importantly, a similar cognitive constraint also applies to information stored in forecasters' memory. In the same way that basing forecasts on all available external information is costly, it is mentally costly to base one's forecasts on all available internal information. Given this constraint, their prior knowledge is imperfectly accessed when they make new projections.

3 The Optimal Mental Representation

We have seen that DM bases her forecasts on two types of information: mental representation of internal information ($m_{i,t}$) and mental representation of external information ($n_{i,t}$). In this section, I discuss the optimal structure of $m_{i,t}$ and $n_{i,t}$.

3.1 The Optimization Problem

The cognitive process is described by the sequence of $\{K_t, \Sigma_{u,t}, \Lambda_t, \Sigma_{\omega,t}\}_{t=0}^{\infty}$. The optimal sequence minimizes the loss function (7) subject to the information environment (3), (4), (5), and (6).

3.2 Optimal Representation of Noisy News $n_{i,t}$

The optimal $n_{i,t}$ is one-dimensional and has the following structure.

Proposition 1. *$\tilde{n}_{i,t}$ is the optimal representation of $N_{i,t}$ such that*

$$\tilde{n}_{i,t} = \kappa_t \cdot E\left[\left. z_t \right| N_t\right] + \tilde{u}_{i,t} \tag{8}$$

for some positive scalar $\kappa_t \in [0, \bar{\kappa}_t]$ and idiosyncratic noise $\tilde{u}_{i,t}$ drawn from $\mathcal{N}\left(0,\, \sigma_{u,t}^2\right)$.

Proof. See Appendix B. □

Intuitively, the optimal representation of $n_{i,t}$ should only capture information in N_t that is useful for predicting z_t. This is because other information in N_t uses up resources but does not further increase the forecast accuracy. Since $\left. z_t \right| N_t$ follows a Gaussian distribution, such information is summarized in the first moment. Therefore, $\tilde{n}_{i,t}$ encodes $E\left[\left. z_t \right| N_t\right]$, which is denoted as follows without loss of generality:

$$E\left[\left. z_t \right| N_t\right] = z_t + \tilde{\nu}_t$$

Here, $\tilde{\nu}_t \sim \mathcal{N}\left(0,\, \sigma_\nu^2\right)$ for some non-negative σ_ν^2 that is taken as given and known to DM.

As one can see from (8), there are combinations of κ_t and $\sigma_{u,t}^2$ that imply the same posterior distribution $\left. z_t \right| m_{i,t}, \tilde{n}_{i,t}$ for any given $m_{i,t}$. Therefore, I impose a normalization so that κ_t alone captures the accuracy of the representation. I assume that

$$Cov\left[\left. z_t, \tilde{n}_{i,t} \right| m_{i,t}\right] = Var\left[\left. \tilde{n}_{i,t} \right| m_{i,t}\right],$$

in which case the posterior uncertainty is determined as

$$\Sigma_{t|t} = \left(1 - \kappa_t\right) \Sigma_{t|t}^m$$

for a given prior uncertainty $\Sigma_{t|t}^m$. That is, observing $\tilde{n}_{i,t}$ reduces the uncertainty about z_t by a factor of $1 - \kappa_t$. This normalization pins down $\sigma_{u,t}^2$ as the following function of κ_t:

$$\sigma_{u,t}^2 = \kappa_t \left(1 - \kappa_t\right) \Sigma_{t|t}^m - \kappa_t^2\, \sigma_\nu^2$$

One can then see that any $\kappa_t \in \left[0, \frac{\Sigma_{t|t}^m}{\Sigma_{t|t}^m + \sigma_\nu^2}\right]$ ensures that the resulting $\sigma_{u,t}^2$ is non-negative.

The value of κ_t is determined by the accuracy constraint (4). Given the optimal structure of $n_{i,t}$, the mutual information between $n_{i,t}$ and N_t equals $\mathcal{I}\left(\tilde{n}_{i,t};\, z_t + \tilde{\nu}_t\right)$. Then, we can pin down κ_t

as a function of ϕ_n:

$$\kappa_t = \frac{\Sigma^m_{t|t}}{\Sigma^m_{t|t} + \frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma^2_\nu\right) + \sigma^2_\nu} \tag{9}$$

We can see that noisier news implies lower κ_t and higher posterior uncertainty. Also note that after long enough learning, the subjective uncertainty $\Sigma^m_{t|t}$ and $\Sigma_{t|t}$ converge to a positive steady-state level for all t. Accordingly, $\kappa_t \to \kappa$.

3.3 Optimal Representation of Noisy Memory $m_{i,t}$

The optimal $m_{i,t}$ is one-dimensional and has the following structure.

Proposition 2. *$\tilde{m}_{i,t}$ is the optimal representation of $(m_{i,t-1}, n_{i,t-1})$ such that*

$$\tilde{m}_{i,t} = \lambda_t \cdot z_{i,t|t-1} + \tilde{\omega}_{i,t+1} \tag{10}$$

for some positive scalar $\lambda_t \in [0,1]$ and idiosyncratic noise $\tilde{\omega}_{i,t}$ drawn from $\mathcal{N}\left(0,\, \sigma^2_{\omega,t}\right)$.

Proof. See Appendix B. □

The intuition for deriving the optimal structure is similar to the derivation of $\tilde{n}_{i,t}$. The optimal representation of $m_{i,t}$ captures information in $(m_{i,t-1}, n_{i,t-1})$ that is useful for predicting z_t. Since $z_t|\, m_{i,t-1}, n_{i,t-1}$ follows a Gaussian distribution, such information is summarized in the first moment. Therefore, $\tilde{m}_{i,t}$ encodes $E\left[z_t|\, m_{i,t-1}, n_{i,t-1}\right]$, which is expressed as $z_{i,t|t-1}$.

As one can see from (10), there are combinations of λ_t and $\sigma^2_{\omega,t}$ that imply the same prior distribution $z_t|\,\tilde{m}_{i,t}$. Therefore, I impose a similar type of normalization assumption as I did for noisy news so that the accuracy of the representation is captured by λ_t alone. I impose the restriction that

$$Cov\left[z_t, \tilde{m}_{i,t}\right] = Var\left[\tilde{m}_{i,t}\right],$$

in which case $Var\left[z_{i,t|t-1}\middle|\,\tilde{m}_{i,t}\right] = (1-\lambda_t)\, Var\left[z_{i,t|t-1}\right]$. That is, observing $\tilde{m}_{i,t}$ reduces the uncertainty about $z_{i,t|t-1}$ by a factor of $1-\lambda_t$. This pins down $\sigma^2_{\omega,t}$ as a function of λ_t in the following form:

$$\sigma^2_{\omega,t} = \lambda_t\left(1-\lambda_t\right) Var\left[z_{i,t|t-1}\right]$$

One can then see that any $\lambda_t \in [0,1]$ ensures that the resulting $\sigma^2_{\omega,t}$ is non-negative.

From the representation structure above, one can see that the forecast accuracy is described by λ_t. Given the posterior uncertainty from the previous period, $\Sigma_{t|t-1}$, the prior uncertainty is determined as follows:

$$\Sigma^m_{t|t} = \Sigma_{t|t-1} + \left(1-\lambda_t\right)\left(Var\left[z_t\right] - \Sigma_{t|t-1}\right)$$

Uncertainty about z_t increases from $\Sigma_{t|t-1}$ to $\Sigma_{t|t}^m$ because prior knowledge is imperfectly represented in the new forecasts.

The value of λ_t is determined by the accuracy constraint (6). Given the optimal structure of $m_{i,t}$, the mutual information between $m_{i,t}$ and $m_{i,t-1}, n_{i,t-1}$ equals $\mathcal{I}\left(\tilde{m}_{i,t};\, z_{i,t|t-1}\right)$. Then, we can pin down λ_t as a function of ϕ_m:

$$\lambda_t = 1 - \phi_m$$

One can see that noisier memory corresponds to lower λ_t and higher prior uncertainty.

4 Cognitive Noise and Biased Forecasts

In this section, I show that forecasts based on the mental representation exhibit forecast biases found in Coibion and Gorodnichenko (2015) and Bordalo, Gennaioli, Ma, and Shleifer (2020b). I illustrate how we can interpret these biases through the proposed model. The model also provides an estimation strategy to infer the extent of cognitive constraints from the survey forecasts.

4.1 Forecasts Subject to Cognitive Constraints

DM's time-t prior belief about z_t is derived as follows:

$$z_{i,t|t}^m = (1-\lambda)\, E\left[z_t\right] + \lambda\, z_{i,t|t-1} + \tilde{\omega}_{i,t}$$

We can see that forecasts are sluggish to incorporate past knowledge because memory is noisy ($\phi_m > 0$). When processing internal information is costly, remembered knowledge about z_t is anchored toward the default prior ($E\left[z_t\right]$). In the case of perfect memory, $z_t|\, m_{i,t}$ equals $z_t|\, m_{i,t-1}, n_{i,t-1}$.

Conditional on this prior belief, the posterior belief evolves according to the following formula:

$$z_{i,t|t} \;=\; (1-\kappa)\, z_{i,t|t}^m + \kappa\, z_t + \kappa\, \tilde{\nu}_t + \tilde{u}_{i,t}$$

We can see that forecasts are sluggish to track the current economy when subject to noisy news. When processing external information is costly, forecasts put less weight on new information and therefore are slow to catch up with new developments in z_t.

Combining these two formulas, beliefs about z_t follow the following law of motion:

$$z_{i,t|t} = (1-\lambda)\,(1-\kappa)\, E\left[z_t\right] + \lambda\,(1-\kappa)\, z_{i,t|t-1} + \kappa\, z_t + (1-\kappa)\, \omega_{i,t} + \kappa\, u_{i,t} \tag{11}$$

The above equation summarizes the features of forecasts subject to cognitive noise. Because of noisy news, DM sluggishly recognizes a change in z_t. Because of noisy memory, DM sluggishly incorporates her past knowledge. And the idiosyncratic cognitive noise from noisy news and noisy memory creates forecast dispersion.

It is helpful to discuss how the noisy-memory assumption changes the predictions of the traditional noisy-information model. If memory is perfect, then beliefs about z_t evolve according to the following formula:

$$z_{i,t|t} = (1 - \kappa^*)\, z_{i,t|t-1} + \kappa^*\, z_t + \kappa^*\, u_{i,t} \tag{12}$$

Comparing (11) to this law of motion, we can see three changes. With noisy memory, (1) prior knowledge receives a smaller weight ($\lambda < 1$), (2) new information receives a bigger weight ($\kappa \leq \kappa^*$), and (3) a new source of cognitive noise appears.

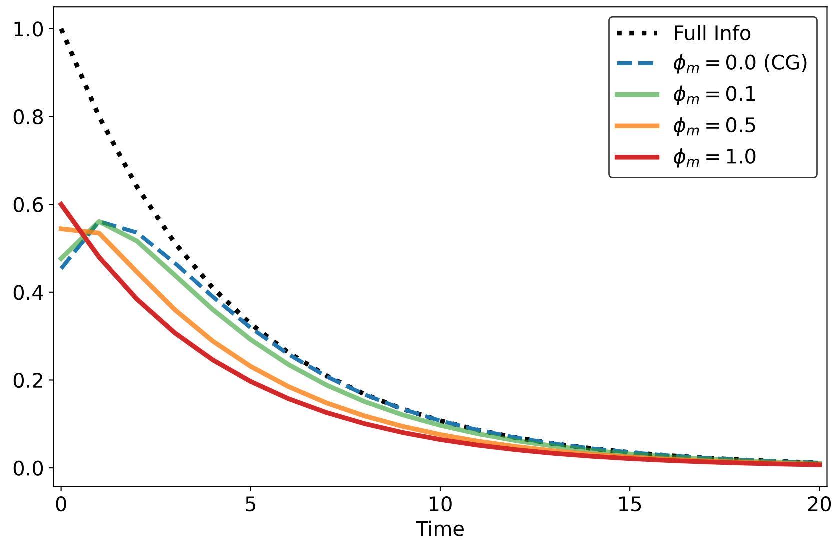
Impulse response function. Figure 1 illustrates the effects of noisy memory when learning about z_t. For this numerical exercise, I use the parameter values $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$ for the data-generating process. I fix the extent of noisy news at $\phi_n = 0.4$.

The top panel shows the impulse response to innovation in z_t. The black dashed line shows the response of z_t. Other lines show the response of forecasts of z_t for varying degrees of noisy memory ϕ_m. The blue line is the perfect-memory case: As DM slowly learns about z_t, her forecasts undershoot the true z_t. With enough learning opportunities, the undershooting disappears, and forecasts closely follow the true y_t. In comparison, the red line is the no-memory case, in which DM has no access to her prior knowledge. Two features stand out. First, the initial response is more significant than the blue line. This is because the Kalman gain is higher when memory is imperfect. And second, learning is slow. Since DM cannot tap into her prior knowledge, learning takes a long time, even with the large Kalman gain. The other colored lines show the in-between cases, and the same intuition applies.

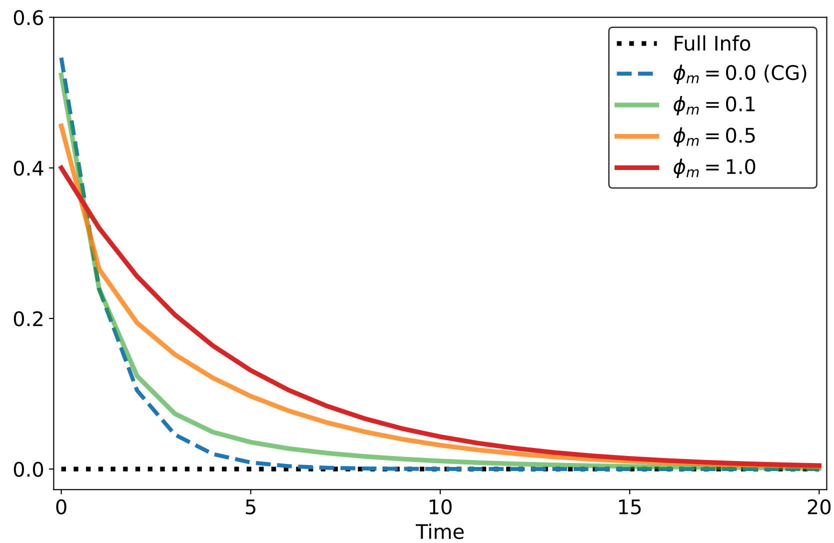
The bottom panel shows the impulse responses of the forecast errors, defined as $z_t - z_{i,t|t}$. When memory is perfect (the blue line), the initially large, positive response diminishes as learning accumulates. When memory is noisy, the forecast errors are initially smaller but remain large even as learning opportunities accrue.

Figure 1: Impulse response of forecasts

(a) Forecast



(b) Forecast errors



The figures show the impulse response to an innovation in z_t. The top panel shows the response of z_t and the forecast of z_t. The bottom panel shows the response of the forecast errors, defined as $z_t - z_{i,t|t}$. The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$. I fix the extent of noisy news as $\phi_n = 0.4$. The black dashed line shows the full-information case of perfect news and memory. Lines with different colors assume a varying degree of noisy memory.

4.2 Biases in Survey Forecasts

In this section, I revisit two regression specifications that test whether survey forecasts deviate from FIRE. Then, I discuss what the test results can inform us about the extent of underlying cognitive noise.

Three building-block assumptions of FIRE. Before investigating the features of survey forecasts, it is helpful to clarify the three assumptions embedded in FIRE. First, forecasters efficiently use all available information at hand. Thus, errors in forecasts are not systematically predictable by any element in the information set. Second, forecasters can access their prior knowledge perfectly. This means that forecast revisions should be in each individual forecaster's information set. Third, forecasters have access to the same complete information.

The first two assumptions predict that an econometrician cannot predict errors that an individual forecaster will make based on the latter's recent forecast revisions. The three assumptions together predict that an econometrician cannot predict errors in the average forecasts based on recent revisions in average forecasts.

Coibion and Gorodnichenko (2015) Regression Specification

Coibion and Gorodnichenko (2015) propose the following regression specification as a joint-hypothesis test for the three FIRE assumptions:

$$y_{t+h} - y_{t+h|t} = \alpha_C + \beta_C \left(y_{t+h|t} - y_{t+h|t-1} \right) + e_{t+h|t} \tag{13}$$

Here, $y_{t+h|t}$ and $y_{t+h|t-1}$ are the average forecasts of $y_{i,t+h|t}$ and $y_{i,t+h|t-1}$.

The authors find a positive β_C for many macroeconomic variables and reject the null hypothesis. They argue that relaxing the full-information assumption can explain the result. Intuitively, if the population does not have access to complete information, revisions in the average forecasts will be sluggish, as at least some people make forecasts based on outdated information. Then, on average, forecasters revise their view about the future sluggishly in response to a change in the economy, and forecast errors are positively correlated with forecast revisions. Furthermore, the authors argue that a larger estimate of β_C can be interpreted as evidence for more significant information frictions.

The expectation-formation model introduced in Section 2 gives new insight into interpreting the regression coefficient.

Proposition 3. *For forecasts subject to cognitive noise, the asymptotic limit of β_C is*

$$\beta_C = \frac{1-\kappa}{\kappa} \left\{ 1 + (1-\lambda) \frac{\lambda (1-\kappa) \rho^2}{1 - \lambda (1-\kappa) \rho^2} \right\}$$

if $\sigma_\nu^2 \to 0$. Furthermore, β_C has the following properties:

1. *$\beta_C > 0$ if $\phi_n > 0$, and $\beta_C = 0$ if $\phi_n \to 0$.*

2. $\frac{\partial \beta_C}{\partial \phi_n} > 0$, *and* $\frac{\partial \beta_C}{\partial \phi_m} < 0$ *if* $\phi_n \leq \bar{\phi}_n \equiv \bar{g}\left(\rho, \sigma_\epsilon^2\right)$.

Proof. See Appendix D. □

Because of noisy news, the proposed model generates a positive β_C. Forecasters update their beliefs sluggishly because they do not have perfect awareness of the current state. As discussed in the previous section, the Kalman gain κ of less than one captures such sluggishness. In addition, noisier news generates a smaller gain and a larger β_C, as argued in Coibion and Gorodnichenko (2015).

A new insight from the proposed model is that noisy memory and noisy news jointly determine the Kalman gain. With noisier memory, the recalled prior knowledge is less accurate. Since uncertainty about the state is higher, forecasters put a larger weight on incoming data, which results in a higher Kalman gain and a lower β_C.

Bordalo, Gennaioli, Ma, and Shleifer (2020b) Regression Specification

Bordalo, Gennaioli, Ma, and Shleifer (2020b) propose the following regression specification as a joint-hypothesis test for the first two FIRE assumptions:

$$y_{t+h} - y_{i,t+h|t} = \alpha_I + \beta_I \left(y_{i,t+h|t} - y_{i,t+h|t-1}\right) + e_{i,t+h|t} \tag{14}$$

The authors reject the null hypothesis and find a negative β_I for many macroeconomic variables in contrast to the result from Coibion and Gorodnichenko (2015). They propose a non-Bayesian expectation model to explain the negative coefficient. The main idea is that forecasters irrationally put too much weight on new observations and over-revise their forecasts. Based on such a model, forecasters are not using available information efficiently, which generates a nonzero β_I. Furthermore, the authors argue that a more negative estimate of β_I can be interpreted as the extent of irrationality.

In contrast, I propose to relax the perfect-memory assumption while keeping the Bayesian-efficiency assumption. The proposed model offers an alternative interpretation of the regression coefficient as follows.

Proposition 4. *For forecasts subject to cognitive noise, the asymptotic limit of* β_I *is*

$$\beta_I = -\frac{(1-\lambda)(1-\kappa)}{2(1-\lambda)(1-\kappa) + \rho^{-2} - 1}$$

if $\rho > 0$. *Furthermore,* β_I *has the following properties.*

1. $\beta_I < 0$ *if* $\phi_m > 0$, *and* $\beta_I = 0$ *if* $\phi_m \to 0$.

2. $\frac{\partial \beta_I}{\partial \phi_n} < 0$, *and* $\frac{\partial \beta_I}{\partial \phi_m} < 0$.

Proof. See Appendix D. □

The regression coefficient captures the bias in underusing past information. Because of noisy memory, forecasts put less weight on past knowledge, which is captured by negative β_I.

Furthermore, noisy news and noisy memory jointly determine this forecast bias. Noisier memory leads to more underuse of past information, which generates a more negative β_I. With noisier news, forecasters rely more on their memory when making forecasts. Since external information is less effective in correcting the bias, β_I is more negative.

Identification of the Extent of Cognitive Constraints

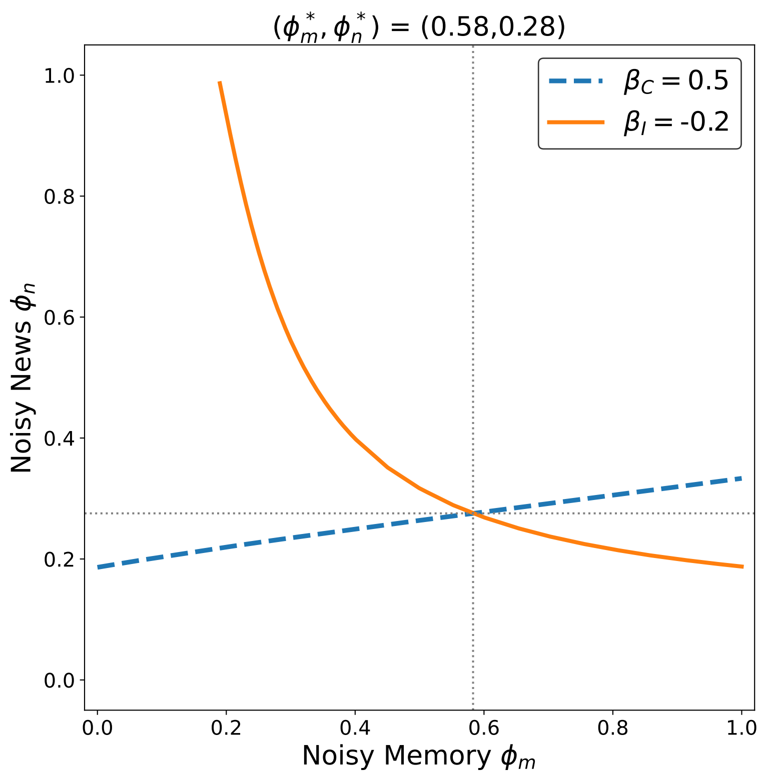
From Propositions 1 and 2, we can see that the two regression coefficients can pin down the severity of noisy news and noisy memory.

Lemma 1. *Given levels of β_C and β_I identify a unique pair of ϕ_n and ϕ_m, if it exists.*

Proof. We can find the pairs of ϕ_n and ϕ_m that generate given levels of β_C and β_I (that is, the iso-curve). The iso-curve for β_C is upward-sloping, and the iso-curve for β_I is downward-sloping. Therefore, if the two iso-curves cross, they only cross once. □

Figure 2 illustrates the lemma. I assume that $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$. The blue solid line is the iso-curve when $\beta_C = 0.5$. And the orange dashed line is the iso-curve when $\beta_I = -0.2$. We can see that the iso-curve for β_C is upward-sloping; more aggressive belief updating due to noisier memory is offset by more sluggish belief updating due to noisier news. We can also see that the iso-curve for β_I is downward-sloping; more underuse of past information due to noisier memory is offset if reliance on memory declines because of less noisy news. These two iso-curves cross once at most, identifying the extent of noisy news and noisy memory that can jointly predict the two estimated regression coefficients.

Figure 2: β_C and β_I jointly identify the extent of cognitive noise



This figure shows the iso-curves for the two regression coefficients in (13) and (14). The blue solid line displays the pairs of noisy-news constraint ϕ_n and noisy-memory constraint ϕ_m that generate $\beta_C = 0.5$. The orange dashed line displays such pairs that generate $\beta_I = -0.2$. The point at which the two lines cross is the estimated extent of noisy news and noisy memory — that is, $\phi_n^* = 0.58$ and $\phi_m^* = 0.28$. The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1$.

5 Extended Model

In Section 2, I assumed that DM is fully aware of the parameters generating z_t. In this section, I assume forecasters are also learning about the long-run mean of the forecast variable. I show that this extension can improve the model predictions in explaining the features of long-run forecasts.

5.1 Learning about the Long Run

Before I estimate the model, I revisit a commonly made assumption in the literature: that people are perfectly aware of the model. It is often motivated by the idea that people adapt to their environment and learn to make optimal economic decisions. However, I show this assumption is not innocuous in the proposed model. As discussed in Azeredo da Silveira et al. (2020), when prior knowledge is imperfectly accessed, forecasters do not reach complete awareness of the model parameters.

One aspect of the environment that is particularly important for making long-horizon forecasts is the mean of the forecast process. Therefore, I assume that DM does not know the exact level of μ and has to learn about it, starting from a Gaussian prior:

$$\mu \sim \mathcal{N}\left(\bar{\mu},\, \Omega\right)$$

The state variable relevant for predicting future realizations is expanded from y_t to (μ, y_t). This is because forecasts for y_{t+h} depend on DM's beliefs about μ and y_t. I denote this state vector as

$$x_t = \begin{pmatrix} \mu \\ z_t \end{pmatrix}.$$

All other assumptions are the same as in Section 2.

5.2 The Optimal Cognitive Process

The optimization problem for deriving the optimal cognitive process is the same as described in Section 2; the optimal process minimizes the objective function (1) subject to the information environment (3), (4), (5), and (6). However, the optimal cognitive process differs from the one introduced in Section 3 because the state variables are multivariate. In this section, I sketch the optimal cognitive process. Detailed derivations are in Appendix C.

Implications of the linear-Gaussian structure. We can see that the initial prior about x_t is Gaussian. Therefore, the linear-Gaussian structure of noisy news and noisy memory again ensures that DM's belief about x_t follows a Gaussian distribution. DM's beliefs about x_t (based on her cognitive state) are described with the following notation:

$$x_\tau|\, m_{i,t} \sim \mathcal{N}\left(x^m_{i,\tau|t},\, \Sigma^m_{\tau|t}\right)$$
$$x_\tau|\, m_{i,t}, n_{i,t} \sim \mathcal{N}\left(x_{i,\tau|t},\, \Sigma_{\tau|t}\right)$$

The loss function. The loss function reduces to

$$\sum_{t=0}^{\infty} \beta^t \, trace\left(\Sigma_{t|t}\, Q\right).$$

Q is a matrix defined as $Q \equiv \sum_{h=1}^{H} \alpha_h \, \alpha_h'$, where $\alpha_h = \begin{pmatrix} 1-\rho^h & \rho^h \end{pmatrix}$.

Optimal representation of noisy news. I first derive the optimal structure of the noisy news $n_{i,t}$. The optimal representation of N_t, denoted as $\tilde{n}_{i,t}$, takes the form

$$\tilde{n}_{i,t} = \tilde{K}_t \cdot E\left[x_t|\, N_t\right] + \tilde{u}_{i,t}$$

for some matrix $\tilde{K}_t$ and idiosyncratic noise $\tilde{u}_{i,t} \sim \mathcal{N}\left(O,\, \Sigma_{u,t}\right)$. The structure is similar to the optimal $n_{i,t}$ in Section 3. Since the forecast accuracy depends on the posterior uncertainty about x_t, the optimal summary of the information in N_t is captured by $E\left[x_t|\, N_t\right]$.

Under the assumed structure of the external news N_t in (2), the optimal $\tilde{K}_t$ and $\sigma_{u,t}$ are determined as follows:

$$\tilde{K}_t = \kappa_t \cdot \frac{\Sigma^m_{t|t}\, e\, e'}{e'\, \Sigma^m_{t|t}\, e}$$
$$\Sigma_{u,t} = \sigma^2_{u,t} \cdot \frac{\Sigma^m_{t|t}\, e\, e'\, \Sigma^m_{t|t}}{\left(e'\, \Sigma^m_{t|t}\, e\right)^{-2}}$$

κ_t and $\sigma^2_{u,t}$ were derived in Section 3. The vector $e' = \begin{pmatrix} 0 & 1 \end{pmatrix}$ picks out z_t from the state vector x_t. The above expression shows that the information represented in $\tilde{n}_{i,t}$ is $E\left[z_t|\, N_t\right]$ (with random errors). This is because the information in N_t about the additional state variable μ is subsumed in $E\left[z_t|\, N_t\right]$.

Optimal representation of noisy memory. The optimal representation of $(m_{i,t-1}, n_{i,t-1})$ is described with $\tilde{m}_{i,t}$ such that

$$\tilde{m}_{i,t} = \tilde{\Lambda}_t \cdot x_{i,t|t-1} + \tilde{\omega}_{i,t}$$

for some matrix $\tilde{\Lambda}_t$ and idiosyncratic noise $\tilde{\omega}_{i,t} \sim \mathcal{N}\left(O,\, \Sigma_{\omega,t}\right)$. Intuitively, it is optimal to represent knowledge about x_t from the internal information $(m_{i,t-1}, n_{i,t-1})$, which is summarized as $E\left[x_t|\, m_{i,t-1}, n_{i,t-1}\right]$. I apply the normalization so that the accuracy of the representation is entirely determined by $\tilde{\Lambda}_t$:

$$Cov\left[x_{i,t|t-1}, \tilde{m}_{i,t}\right] = Var\left[\tilde{m}_{i,t}\right]$$

This pins down the memory-noise variance $\Sigma_{\omega,t}$ as a function of $\tilde{\Lambda}_t$.

$$\Sigma_{\omega,t} = \left(I - \tilde{\Lambda}_t\right) Var\left[x_{t|t-1}\right] \tilde{\Lambda}_t'$$

Any $\tilde{\Lambda}_t$ is feasible as long as the resulting $\Sigma_{\omega,t}$ is a proper variance-covariance matrix (that is, symmetric and positive-semidefinite).

In the appendix, I describe how $\tilde{\Lambda}_t$ can be derived. The complication arises because the information constraint (6) cannot completely determine the noisy memory anymore. To see why, note that the constraint reduces to

$$\mathcal{I}\left(m_{i,t};\, m_{i,t-1}, n_{i,t-1}\right) = -\frac{1}{2}\det\left(I - \tilde{\Lambda}_t\right) \leq -\frac{1}{2}\ln\phi_n.$$

That is, this constraint limits the determinant of $I - \tilde{\Lambda}_t$, leaving the elements of $\tilde{\Lambda}_t$ to be specified.

When solving for $\tilde{\Lambda}_t$, I consider a myopic case, in which $\beta \to 0$. I first define a matrix Γ_t that is crucial for determining the $\tilde{\Lambda}_t$. I call this matrix a *memory-priority* matrix and define it as follows:

$$\Gamma_t = \left(I - K_{t+1}\right)' Q \left(I - K_{t+1}\right)$$

The matrix Γ_t roughly captures how some information receives higher priority than other information. Two matrices show why I make such interpretations. First, the elements in $(I - K_{t+1})$ would be large if external information oes not resolve much uncertainty about the state, in which case a more accurate memory would be helpful. Second, the matrix Q is from the loss function of incorrect forecasting. If some elements in Q were high, more accurate memory would be helpful.

I show in the appendix that $Var\left[x_{i,t+1|t}\right]^{\frac{1}{2}}\ \Gamma_t\, Var\left[x_{i,t+1|t}\right]^{\frac{1}{2}}$ can be eigen-decomposed to $U_t\, G_t\, U_t'$, where U_t is an orthonormal matrix storing the eigenvectors and G_t is a diagonal matrix storing eigenvalues in descending order (that is, $g_{1,t} > g_{2,t}$). Then, the optimal Λ_t satisfies

$$\tilde{\Lambda}_t = Var\left[x_{i,t+1|t}\right]^{\frac{1}{2}}\ U_t\, D_t\, U_t'\, Var\left[x_{i,t+1|t}\right]^{-\frac{1}{2}},$$

where a diagonal matrix D_t is defined to be

$$D_t = \begin{cases} \begin{pmatrix} 1 - \left(\frac{g_{2,t}}{g_{1,t}}\phi_m\right)^{\frac{1}{2}} & 0 \\ 0 & 1 - \left(\frac{g_{1,t}}{g_{2,t}}\phi_m\right)^{\frac{1}{2}} \end{pmatrix} & \text{if} \quad \phi_m < \frac{g_{2,t}}{g_{1,t}} \\ \\ \begin{pmatrix} 1 - \phi_m & 0 \\ 0 & 0 \end{pmatrix} & \text{otherwise.} \end{cases}$$

One can easily see that $\det\left(I - \Lambda_t\right) = \phi_m$. The derivation above shows how the rank of the memory variable is determined. The first case is when the dimension of the remembered knowledge is not reduced (that is, $m_{i,t+1}$ is two-dimensional). In this case, the first diagonal element in D_t is higher than the second one, indicating that the corresponding orthogonalized factor receives a higher weight. The second case is when memory stores information in $x_{i,t+1|t}$ in a lower dimension. The first diagonal element in D_t receives the biggest possible weight satisfying the memory constraint, while the second element is zero.

Summary. We have seen the derivation for optimal noisy news and noisy memory, which is described by the sequence of $\{K_t, \Sigma_{u,t}, \Lambda_t, \Sigma_{\omega,t}\}_{t=0}^{\infty}$. The time-$t$ prior belief is described with $\tilde{\Lambda}_t$:

$$x_{t|t}^m = x_{t|t-1} + \left(I - \tilde{\Lambda}_t\right)\left(E\left[x_t\right] - x_{t|t-1}\right) + \tilde{\omega}_{i,t}$$
$$\Sigma_{t|t}^m = \Sigma_{t|t-1} + \left(I - \tilde{\Lambda}_t\right)\left(Var\left[x_t\right] - \Sigma_{t|t-1}\right)$$

And the posterior belief is described with $\tilde{K}_t$:

$$x_{i,t|t} = \left(I - \tilde{K}_t\right) x_{i,t|t}^m + \tilde{K}_t\, x_t + \tilde{\nu}_t + \tilde{u}_{i,t}$$
$$\Sigma_{t|t} = \left(I - \tilde{K}_t\right) \Sigma_{t|t}^m$$

Here, $\tilde{\nu}_t \sim \mathcal{N}\left(O, \Sigma_\nu\right)$, whose variance is defined as $\Sigma_\nu = \kappa_t^2 \left(e' \, \Sigma_{t|t}^m \, e\right)^{-2} \Sigma_{t|t}^m \, e \, e' \, \Sigma_{t|t}^m$.

5.3 Perpetual Uncertainty about the Long Run

This section briefly discusses how DM learns about the long-run mean when she is subject to cognitive noise. Based on this discussion, I show the model predictions about the forecast-error-revision test for different forecast horizons.

When DM can access her internal information perfectly, she has complete access to all the past noisy news. In this case, the subjective uncertainty about the mean is

$$Var\left[\mu \,|\, n_{i,t}, n_{i,t-1}, \cdots, n_{i,0}\right] = \left(\Omega^{-1} + t \times c\right)^{-1},$$

where Ω is the prior variance about μ, and c is a constant. We can see that the precision of knowledge linearly increases in time; the uncertainty eventually converges to zero after a long learning period.

Noisy memory qualitatively changes this prediction as investigated in Azeredo da Silveira, Sung, and Woodford (2020). If DM imperfectly accesses internal information, $Var\left[\mu \,|\, m_{i,t}, n_{i,t}\right]$ does not converge to zero even after a long learning period. The intuition is straightforward: cognitive noise prevents forecasters from reaching complete awareness even after an infinitely long learning period.

Why does it matter that DM is imperfectly aware of the long-run mean? It matters because DM will continuously update her beliefs about the mean as new data come, although she correctly understands that the mean is a constant parameter. When y_t is high, the DM partly attributes it to higher-than-expected μ and expects future y_t to be persistently high. This prediction is similar to extrapolative-expectation models in the finance literature. My model implies that a limited memory might be the reason such extrapolation occurs.

Impulse response function. Figure 3 illustrates the effect of learning about the long run. I use the same data-generating process as Figure 1 and set the cognitive parameters as $\phi_n = 0.4$, $\phi_m = 0.1$, and $\Omega = 1$.

The top panel shows the impulse response to innovation in z_t. The black dashed line is the response of z_t. The blue line is the response of forecasts for z_t. As in Figure 1, learning about y_t is sluggish because of noisy news.[3] The orange line shows the forecast for μ. As discussed earlier, DM perceives that z_t is high partly because the long-run mean is high and revises her belief about μ upward.

The bottom panel of Figure 3 displays the response of four-quarter-ahead forecasts for varying degrees of Ω. I realign the lines to compare forecasts to the realized z_{t+4}. We can see whether forecasts undershoot or overshoot compared to the black dashed line. We see initial undershooting for all values of Ω because of the noisy news. However, forecasts start overshooting after a few periods for some Ω. When Ω is high, DM revises her beliefs about the long-run mean too much, which offsets the undershooting due to noisy news. In this case, the forecast errors, defined as $z_{t+4} - z_{i,t+4|t}$, are initially positive in response to innovation in z_t but soon turn negative. This prediction is consistent with findings in Angeletos et al. (2021). The authors analyze the professional forecasters' year-ahead forecasts for unemployment and inflation and their impulse response to a specific shock series constructed by Angeletos et al. (2020).

Error-revision regression. The perpetual uncertainty about the long run also implies that the regression coefficients in the forecast error-revision test (13) and (14) will not be constant for different forecast horizons.

Consider the regression coefficient applied to forecasts for μ. Denoting the mean forecasts as $\hat{\mu}_{i,t} \equiv E\left[\mu \middle| m_{i,t}, n_{i,t}\right]$ and the average forecasts as $\hat{\mu}_t \equiv \int \hat{\mu}_{i,t} d\, i$, we can see that

$$\beta_C^{\mu} = \frac{Cov\left[\mu - \hat{\mu}_t, \hat{\mu}_t - \hat{\mu}_{t-1} \middle| \mu\right]}{Var\left[\hat{\mu}_t - \hat{\mu}_{t-1} \middle| \mu\right]} = -\frac{1}{2}$$
$$\beta_I^{\mu} = \frac{Cov\left[\mu - \hat{\mu}_{i,t}, \hat{\mu}_{i,t} - \hat{\mu}_{i,t-1} \middle| \mu\right]}{Var\left[\hat{\mu}_{i,t} - \hat{\mu}_{i,t-1} \middle| \mu\right]} = -\frac{1}{2}.$$

The derivation is straightforward. We can deduce that $\beta_C = -\frac{Var[\hat{\mu}_t|\mu] - Cov[\hat{\mu}_t, \hat{\mu}_{t-1}|\mu]}{2(Var[\hat{\mu}_t|\mu] - Cov[\hat{\mu}_t, \hat{\mu}_{t-1}|\mu])}$ and must equal $-\frac{1}{2}$. The same reasoning applies to β_I.[4] Forecasters revise their views about μ although μ is a fixed parameter.

[3] We can see that the impulse response of z_t more closely tracks z_t in Figure 3 than in Figure 1. This is because uncertainty about the long run increases uncertainty about z_t, pushing up the Kalman gain.

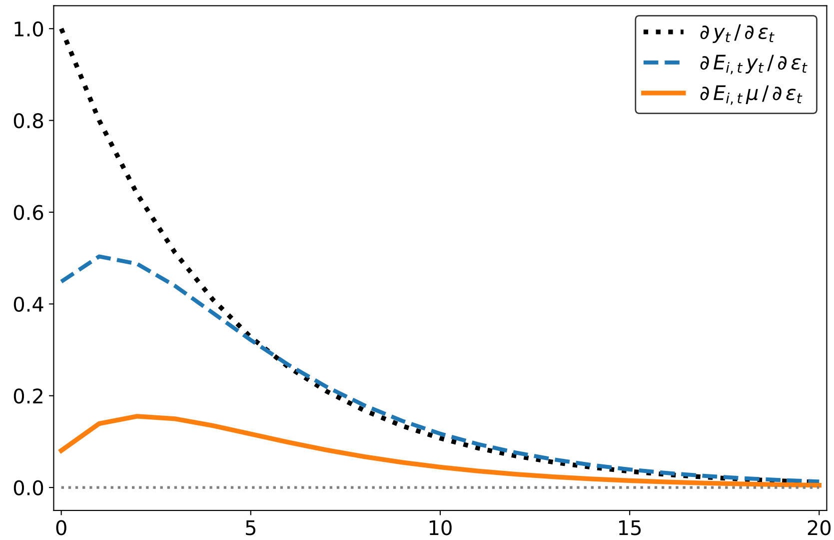
[4] Derivations for other horizons are in Appendix F.

Figure 4 illustrates the model predictions for β_C and β_I for varying forecast horizons. I fix the degree of noisy news and noisy memory at levels in Figure 2 that generate the targeted β_C and β_I. I use $\Omega = 0.2$; this level corresponds to the posterior variance of μ if DM had access to twenty years of data. The figure shows that both coefficients become more negative for longer forecast horizons. As shown earlier, for forecasts far enough ahead, β_C^μ and β_I^μ are close to $-\frac{1}{2}$.

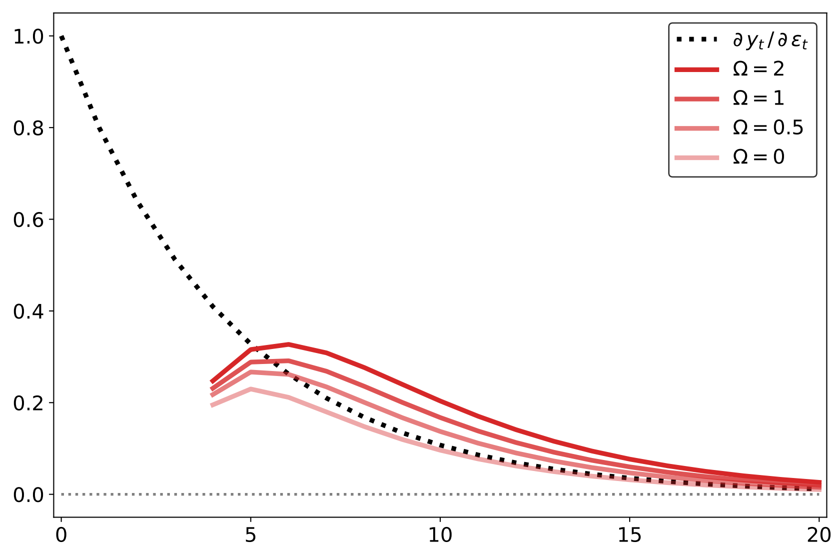
The pattern in Figure 4 is in line with empirical findings in the literature. d'Arienzo (2020) and Wang (2021) analyze professional forecasters' projections of interest rates. Both authors find that longer-horizon forecasts feature more negative biases when the regressions (13) and (14) are estimated. Bordalo, Gennaioli, Porta, and Shleifer (2019) and Bordalo, Gennaioli, La Porta, and Shleifer (2020a) find a similar pattern for stock analysts' forecasts for companies' long-term earnings.

Figure 3: Impulse-response functions when learning about the long run

(a) Forecasts of z_t and μ

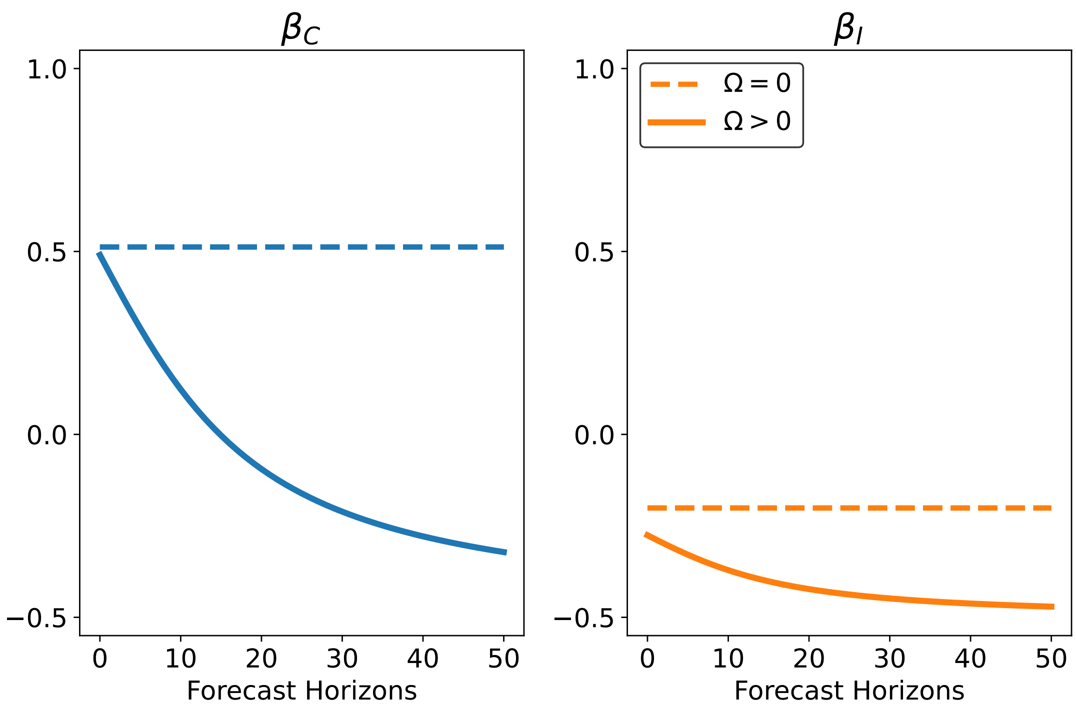


(b) Forecasts of z_{t+4}



The figures show the impulse response to an innovation in y_t. The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1$. The top panel shows the response of y_t and the forecast of y_t and μ. I fix the cognitive noise as $\phi_n = 0.4$ and $\phi_m = 0.1$, and I set $\Omega = 1$. The bottom panel shows the response of four-period-ahead forecasts ($y_{i,t+4|t}$). Different lines assume varying degrees of Ω, the initial uncertainty about μ.

Figure 4: β_C and β_I when learning about the long run



This figure shows model predictions of the two regression coefficients in (13) and (14) for different forecast horizons. The extent of cognitive noise is from Figure 2: $\phi_n^* = 0.28$ and $\phi_m^* = 0.58$. The gray solid line is the model prediction when DM does not have to learn about the long run ($\Omega = 0$). The black dashed line is when DM learns about the long run ($\Omega = 1$). The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1$.

6 Estimating the Extent of the Cognitive Constraints

In this section, I estimate the two cognitive constraints by using professional forecasters' survey data.

6.1 Data

Survey forecast data are from the Survey of Professional Forecasters (SPF), administered by the Federal Reserve Bank of Philadelphia. Once every quarter, around forty forecasters (mostly from academia and banks) participate in this survey. The earliest survey started in 1968. I use survey forecasts made until the second quarter of 2022.

Among the survey questions, those in the section titled "The U.S. Business Indicators" ask forecasters to submit their views about aspects of the overall US economy, which include output, price level, labor and housing markets, and cost of borrowing. I investigate whether the proposed model can explain features of survey forecasts made for that section.[5] Table 1 lists the variables.

For data on the time series of macroeconomic variables, I use the Real-Time Data Set from the Federal Reserve Bank of Philadelphia whenever possible. This data set provides the history of data releases for each variable. Since the variables in the National Income and Product Accounts are often redefined or reclassified, the final data release (that is, the most recently available data) often does not include the same variables forecast by the professional forecasters in the data set. Therefore, I compare the initial releases of each variable to the corresponding SPF forecasts.

6.2 Estimation Strategy

I estimate four parameters that affect how DM makes forecasts about the macroeconomic variables: ϕ_n and ϕ_m (the severity of the two cognitive limitations), σ_ν^2 (the amount of correlated noise), and Ω (the unconditional prior uncertainty about the long-run mean).[6] The parameters describing the data-generating process are estimated from the realized macroeconomic variables. I assume that each variable is described as a univariate autoregressive process. Related parameters are in Appendix E. Finally, I assume that the longest forecast horizon of the loss function (1) is eight quarters ahead since the SPF asks forecasters to submit their forecasts for up to two years ahead for the "The U.S. Business Indicators" section.

I transform the survey forecast data so that the unit of forecasts is the log difference from the previous quarter for most variables. I use change from the previous quarter for the unemployment rate and the three financial variables in Table 1. Surveyed forecasters make projections for different horizons, so all forecasts are annualized to make the units consistent. I use forecasts up to four quarters ahead.

[5] There have been some categorical changes, as the survey forms changed over time, but I include eleven variables that are consistently included most of the time.

[6] Failure to consider the correlated noise ν_t can bias the model estimation. This is because the estimated regression coefficient β_C from (13) is attenuated when forecast noise is correlated among forecasters (Coibion and Gorodnichenko (2015) and Gemmi and Valchev (2021)).

I drop some observations to restrict the influence of a few outlier variables. In each period, I remove forecasts if they are five quantiles outside the median level. I remove forecasters if they participate for fewer than ten periods. I further restrict samples to measure the forecast behavior in the normal business cycle. During periods of big swings in the macroeconomy such as the COVID-19 pandemic, it is likely that forecasters use different forecasting methods and therefore exhibit different behaviors. Since my model does not capture such structural changes, I use a simple algorithm to remove likely structural-change episodes. Namely, I compute the average size of forecast revisions among forecasters each period and remove the top five percentile periods. This procedure systematically identifies significant revision episodes, removing the beginning of the pandemic for unemployment but not for less affected variables.

Estimation targets. The first two data moments I use are the regression coefficients described earlier: β_C from (13) and β_I from (14). The forecast error-revision pair is available for the forecast horizon for up to three quarters. I estimate the regression by pooling the four forecast horizons.[7] For the individual-level regression, I include individual and horizon dummies to purge variations due to the fixed effects.

I panel-bootstrap the SPF individual-forecast data and build bootstrap samples of the targeted moments. Each sample contains on average forty individual forecasters, as in the survey data. The first two panels of Table 1 report this coefficient. The table reports the median and confidence interval of 5%–95% estimates. As discussed in Section 4.2, we see positive β_C and negative β_I across the variables. These two moments can identify the underlying degree of information constraints, given the two remaining parameters σ_ν^2 and Ω. I also report the OLS estimates in Table 3. The bootstrapped estimates and the OLS estimates are similar.

Two more moments are used to estimate the model. These moments are informative about σ_ν^2 and Ω. Based on Gemmi and Valchev (2021), I measure the size of Kalman gains using the following specification:

$$\left(y_{i,t+h|t} - y_{i,t+h|t-1}\right) - \left(y_{t+h|t} - y_{t+h|t-1}\right) = \alpha_K + \beta_K \left(y_{t+h|t-1} - y_{i,t+h|t-1}\right) + error_{i,t+h|t-1} \quad (15)$$

This specification estimates how forecasters revise their views about the current economy in response to news about it. The strategy is to partial out the effects of the correlated noise by demeaning individual forecasts. Since the correlated noise attenuates β_C, comparing the above regression coefficient to the Kalman gain implied by β_C is informative about the degree of correlated noise.[8]

I pool the forecast horizons and control for individual-forecaster and forecast-horizon fixed

[7] I pool the different forecast horizons for two reasons. An obvious reason is to increase power. But more importantly, I am interested in estimating the constraints in processing information about the near-term economy, not just the current economy. A literal interpretation of the model is that DM gets news only about the current economy. (Since the time unit is a quarter, DM gets news about the current quarter only.) However, it would be realistic to assume that forecasters learn about the near-term economy.

[8] The authors show that the new estimate of the Kalman gain is smaller for most macroeconomic variables they study.

effects. The right panel of Table 1 reports the regression coefficient. The table reports the median and confidence interval of 5%–95% estimates, and the OLS estimates are in Table 3.

I use a similar specification to measure how long-term forecasts are revised in response to news about the near-term economy. We need frequent long-term forecast data to estimate this regression. The SPF collects these data for the Consumer Price Index (CPI) but not for other macroeconomic variables. I use forecasts for the annual average rate of headline CPI inflation over the next ten years to estimate the regression. The coefficient is estimated to be 0.0862, statistically significant at the 1% level, with a standard error of 0.0175. More details are in Appendix E. Since data are not available to conduct a similar analysis for other macroeconomic variables, I target the estimated coefficient for all variables. While it is not feasible to verify the validity of this assumption, we can at least see that the estimated regression coefficient for (15) is broadly similar across variables.

6.3 Estimation Results

I now estimate parameters that fit each bootstrapped sample discussed in the previous section. I report the median estimate and the 5%–95% confidence band in Table 4. Figure 6 reports the estimates of noisy news ϕ_n and noisy memory ϕ_m.

In Section 4, I showed that the methodology in Coibion and Gorodnichenko (2015) underestimates the magnitude of ϕ_n because it misattributes the extra sensitivity from noisy memory to low ϕ_n. To investigate the extent of underestimation, I repeat the estimation procedure while assuming $\phi_m = 0$. In this case, I estimate two parameters, ϕ_n and σ_ν^2, that match the two estimation targets, β_C and β_K.

The top panel in Figure 6 compares ϕ_n estimated using the proposed model to that estimated assuming perfect memory. As expected, the estimated ϕ_n is larger with noisy memory. On average, the baseline ϕ_n is twice as large as ϕ_n estimated the using Coibion and Gorodnichenko (2015) methodology. The bottom panel illustrates the estimated ϕ_m. For most variables, ϕ_m is significant and positive. Overall, the estimated parameters are somewhat stable: the average levels are $\phi_n = 0.31$ and $\phi_m = 0.24$; the median levels are $\phi_n = 0.34$ and $\phi_m = 0.22$.

Table 2 assesses the model fit using the point estimate. The top and bottom panels show the targeted and untargeted moments, respectively. This table reports the average levels across macroeconomic variables.

We confirm that the model matches the targeted moments well. For untargeted moments, I show variations in forecasts and forecast revisions. For each variable, I report variations in the time series (that is, dispersion of the consensus forecasts) and in the cross section (that is, dispersion of the individual forecasts at any given time). All measures are the standard deviation scaled by the standard deviation of the forecast variable. We can see that the estimated model has a reasonable quantitative fit.

Figure 5 illustrates the fit of untargeted moments for all macroeconomic variables. Although the model is too stylized to replicate variations across macroeconomic variables perfectly, it generates a good fit. The detailed data for this figure are available in Table 6.

Table 1: Estimated regression coefficients

	β_C	CI	β_I	CI	β_K	CI
Nominal Gross Domestic Product	0.55	(0.43,0.66)	-0.27	(-0.3,-0.24)	0.55	(0.52,0.58)
Real Gross Domestic Product	0.36	(0.26,0.45)	-0.24	(-0.27,-0.21)	0.61	(0.58,0.63)
GDP Chain-Weighted Price Index	0.56	(0.43,0.69)	-0.32	(-0.36,-0.28)	0.6	(0.57,0.64)
Corporate Profits after Taxes	0.49	(0.32,0.66)	-0.44	(-0.48,-0.4)	0.51	(0.49,0.53)
Civilian Unemployment Rate	0.56	(0.51,0.62)	-0.05	(-0.08,-0.02)	0.63	(0.6,0.65)
Industrial Production Index	0.53	(0.44,0.61)	-0.18	(-0.22,-0.15)	0.57	(0.55,0.61)
Housing Starts	0.41	(0.31,0.49)	-0.27	(-0.32,-0.22)	0.58	(0.55,0.6)
Consumer Price Index	0.46	(0.32,0.61)	-0.17	(-0.22,-0.12)	0.57	(0.53,0.6)
Treasury Bill Rate, 3-month	0.28	(0.2,0.34)	-0.01	(-0.03,0.01)	0.73	(0.69,0.77)
AAA Corporate Bond Yield	0.03	(-0.03,0.09)	-0.35	(-0.38,-0.32)	0.68	(0.66,0.7)
Treasury Bond Rate, 10-year	0.26	(0.21,0.33)	-0.12	(-0.15,-0.1)	0.7	(0.67,0.73)

Figure 5: Estimated β_C and β_I

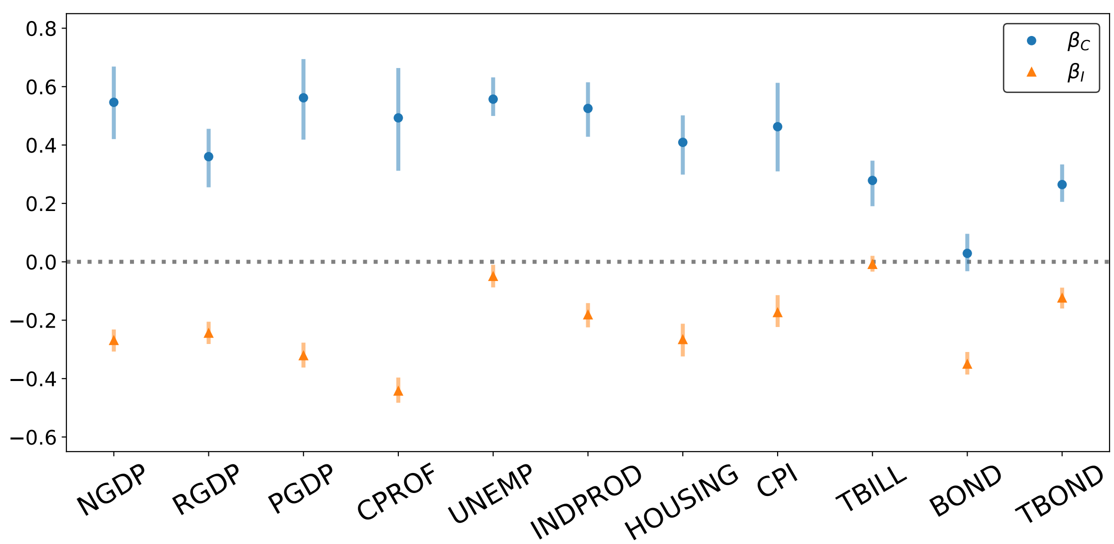
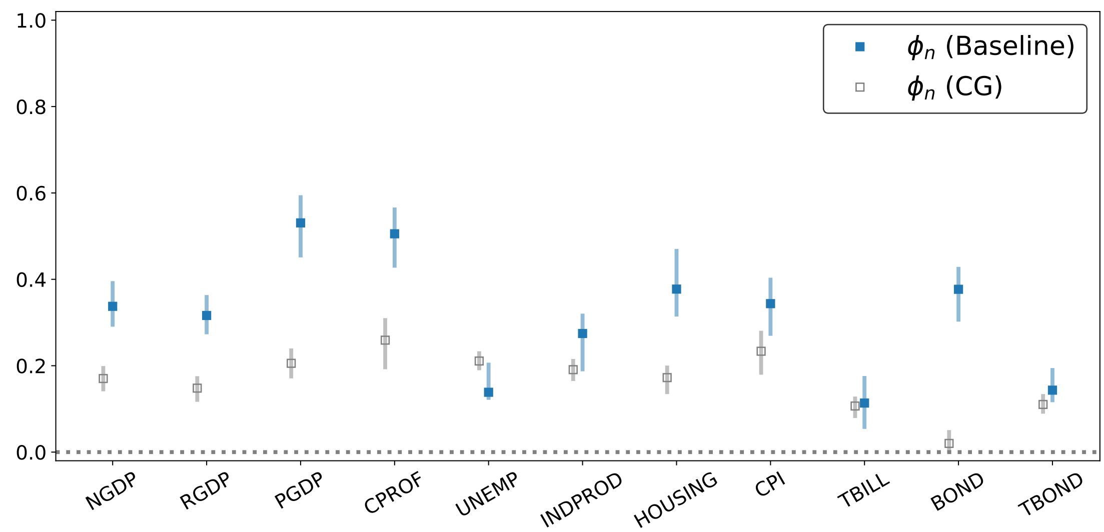


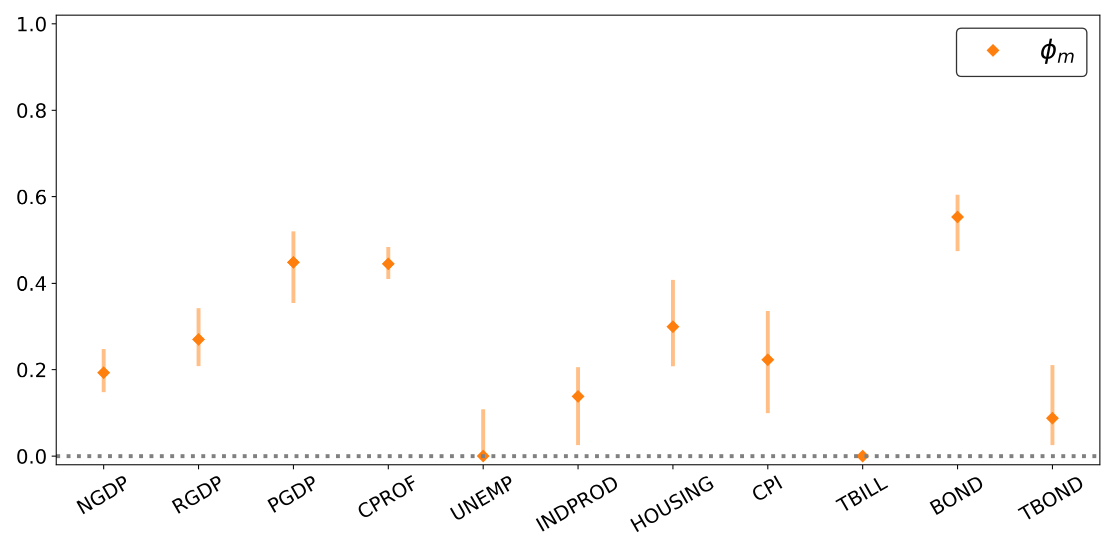
Table 1 reports the estimated regression coefficients. I study the variables in the SPF's "U.S. Business Indicators" section. From left to right, each panel presents the coefficients in (13), (14), and (15). The last two regressions include individual and horizon fixed effects. I panel-bootstrap the SPF data. The dot is the median estimate, and the error band shows the 5% and 95% estimates. The OLS estimates are reported in Table 3. Figure 5 visualizes the estimation of β_C and β_I, whose x-axis uses the abbreviated variable names that are in the same order as in the table above.

Figure 6: Estimated Parameters

(a) Noisy News



(b) Noisy Memory



This figure's full variable names are in Table 1. The top panel reports the estimated extent of noisy news (ϕ_n), and the bottom panel reports that of noisy memory (ϕ_m). Estimation targets the panel-bootstrapped moments discussed in Table 1. The dot is the median estimate, and the error band contains the 5% and 95% estimates. Table 4 reports the detailed numerical results. In the top panel, I compare the estimated ϕ_n (labeled as "Baseline") to the estimation achieved under the Coibion and Gorodnichenko (2015) assumption (labeled as "CG"). For the latter, I impose $\phi_m = 0$ and estimate two parameters (ϕ_n and σ_ν^2) that match two targets (β_C and β_K).

Table 2: Model Fit

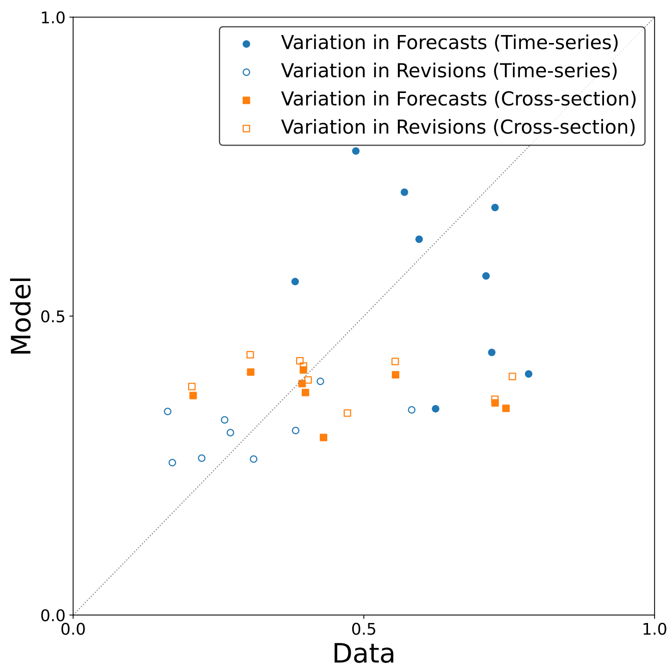
(a) Targeted moments (average across macroeconomic variables)

β_C		β_I		β_K		$\beta_{\mu,K}$	
Data	Model	Data	Model	Data	Model	Data	Model
0.41	0.41	-0.22	-0.2	0.61	0.61	0.08	0.07

(b) Not-targeted moments (average across macroeconomic variables)

Variation in Forecasts				Variation in Revisions			
Time Series		Cross Section		Time Series		Cross Section	
Data	Model	Data	Model	Data	Model	Data	Model
0.6	0.6	0.43	0.36	0.31	0.32	0.44	0.38

Figure 7: Not-targeted moments (all macroeconomic variables)



The tables evaluate the fit of the estimated model when using the median estimates in Table 4. The upper panel shows the targeted moments, and the lower panel shows untargeted moments. Both panels report the average value across all macroeconomic variables in Table 1. For untargeted moments, I report variations of forecasts and forecast revisions in the time series and cross sections, averaged across four consecutive forecast horizons (current to three quarters ahead). The unit of all measures is the standard deviation scaled by the standard deviation of the forecast variables. The figure illustrates the untargeted moments for all macroeconomic variables. The detailed data for this figure are available in Table 6.

7 An Illustrative Macroeconomic Model

In this section, I study the macroeconomic implications of the proposed expectation-formation model. Using a standard New Keynesian model, I show that if expectations are formed as in my model, inflation may be more variable, worsening the central bank's policy trade-off in stabilizing inflation and output. I discuss the efficient monetary policy in this environment and what harm can be done if the central bank conducts monetary policy that is only efficient under conventional expectation assumptions.

7.1 Firms' Decision Problem

Optimal Price Setting

Suppose firm i reconsiders its price $P_{i,t}$ in period t. The new price that it chooses maximizes the expected value of the firm's (current market value) profits. This pricing decision does not constrain any future decisions. Thus, it suffices to consider the effects of the choice on expected profits in those future states in which the price has not yet again been re-optimized. The firm's new price solves the following problem:

$$\max_{P_{i,t}} \quad E_{i,t}\left[\sum_{h=0}^{\infty} \alpha^h \, Q_{t,t+h}\left(P_{i,t}\, Y_{i,t+h|t} - \Psi_{t+h}\left(Y_{i,t+h|t}\right)\right)\right]$$

Here, α is the probability of not resetting prices, $Q_{t,t+h}$ is the stochastic discount factor for evaluating the future nominal payoffs generated at $t+h$, $Y_{i,t+h|t}$ is the output demanded in period $t+h$ if the price remains at the one chosen at time t, and Ψ_{t+h} is the (nominal) cost function at time $t+h$. Firm i takes into account that the demand $Y_{i,t+h|t}$ is given as

$$Y_{i,t+h|t} = \left(\frac{P_{i,t}}{P_{t+h}}\right)^{\eta} C_{t+h},$$

where η is the elasticity of substitution among goods, P_{t+k} is the aggregate price at time $t+h$, and C_{t+h} is the aggregate consumption at time $t+h$.

I use the notation $E_{i,t}$ to denote firm i's subjective expectation at time t. While in the conventional New Keynesian model $E_{i,t}$ refers to full-information rational expectations, I propose that the firm's expectations are formed according to the cognitive limitations proposed in earlier sections. The firm's objective depends only on aggregate conditions at the various dates $t+h$. Thus, under rational expectations, the optimal price $P^*_{i,t}$ would be the same for all i that reconsider their price at date t. However, under the expectation-formation model proposed in this paper, the optimal choice $P^*_{i,t}$ may differ across firms because of their differing expectations.

The firm's optimal price $P^*_{i,t}$ is derived using the first-order condition. Below I describe the first-order Taylor expansion of this condition around the zero-inflation steady state (I use lowercase to denote the log of the variable denoted in uppercase):

$$p_{i,t}^{*}-p_{i,t-1}=E_{i,t}\left[\sum_{h=0}^{\infty}(\alpha\beta)^{h}\left\{(1-\alpha\beta)\left(mc_{t+h}-mc\right)+\pi_{t+h}\right\}\right]$$

Here, mc_{t+h} is the log of real marginal cost at $t+h$ (mc is its steady-state value), and π_{t+h} is inflation at $t+h$ defined as $\log P_{t+h}-\log P_{t+h-1}$. As detailed in Appendix G, the marginal costs do not depend on the quantity that a firm supplies. This is because of the assumed feature of the production function that the marginal product of labor does not depend on the quantity of production. Thus, firm i treats the nominal marginal costs as evolving independently of its own pricing decision; they only depend on aggregate variables that the firm takes as given.[9] Let us define

$$z_{t+h}\equiv(1-\alpha\beta)\left(mc_{t+h}-mc\right)+\pi_{t+h}. \tag{16}$$

Thus, the firm's expectations of the current and future z_t determine its subjectively optimal price:

$$p_{i,t}^{*}-p_{i,t-1}=E_{i,t}\left[\sum_{h=0}^{\infty}(\alpha\beta)^{h}z_{t+h}\right] \tag{17}$$

7.2 Aggregate Economy

Real Marginal Costs

The real marginal costs are derived from the rest of the economy. As detailed in Appendix G, the household optimization problem and market-clearing conditions imply that

$$mc_t-mc=\chi\,x_t+e_t. \tag{18}$$

χ depends on the elasticities of the consumption and labor utility functions, and x_t is defined as $y_t-y_t^e$, where y_t^e is the efficient level of output. Finally, e_t is the cost-push shock. While I do not take a stance on the source of cost-push shocks, one example is a time-varying, exogenous wage markup. The cost-push shock is a transitory i.i.d. shock fluctuating around zero.

Monetary Policy

Because of cost-push shocks, it is infeasible for the central bank to stabilize both inflation and the output gap fully. Thus, the central bank faces a policy trade-off in stabilizing the two variables. I assume that monetary policy is specified by a targeting rule of the form

$$x_t=-s\,\pi_t, \tag{19}$$

[9] I introduce this assumption for the sake of simplicity. However, even when the firm's marginal product of labor varies with the quantity supplied, the subjectively optimal price will still depend only on its expectations about aggregate economic variables. See Gali (2008, Chapter 3).

where s is a constant scalar that I later calibrate to match the relative variability of the output gap to inflation in the data.

The targeting rule illustrates the relationship between x_t and π_t that the central bank seeks to maintain in response to a fluctuation in the economy. The rule implies that the central bank accepts inflation higher than its long-run target (assumed to be zero in the model) if and only if there is a negative output gap. Likewise, the targeting rule requires inflation to be lower than the long-run target when there is a positive output gap at the same time. The implication of such a targeting rule for the path of interest rates can be derived using the household intertemporal optimization condition.

Aggregation

Once firms reconsider their price and choose their subjectively optimal price $P_{i,t}^*$, the aggregate price index is formed according to

$$P_t = \left[\alpha \left(P_{t-1}\right)^{1-\eta} + (1-\alpha)\left(P_t^*\right)^{1-\eta}\right]^{\frac{1}{1-\eta}},$$

where $P_t^* \equiv \int P_{i,t}^* \, di$ is the average reset price of firms that reconsider their prices at time t. The first-order Taylor expansion of the price index implies $\pi_t = (1-\alpha)\left(p_t^* - p_{t-1}\right)$. Therefore, we can derive the aggregate inflation by averaging the expectations of the different firms:

$$\pi_t = (1-\alpha)\bar{E}_t \left[\sum_{h=0}^{\infty} (\alpha\beta)^h z_{t+h}\right] \tag{20}$$

Here, $\bar{E}_t$ averages the expectations $E_{i,t}$ of all individual firms.

Determination of z_t

By substituting (18) and (19) into (16), we can deduce that z_t is determined as follows:

$$z_t = \{1 - (1-\alpha\beta)\,\sigma\, s\}\,\pi_t + (1-\alpha\beta)\, e_t \tag{21}$$

Equations (20) and (21) together imply that z_t is determined by firms' expectations about current and future z_t and the exogenous shock e_t. Thus, once we specify how firms forecast z_t, we have a complete theory of how inflation, the output gap, and z_t evolve.

7.3 Firms' Macroeconomic Expectations

Suppose that firms form their forecasts under the assumption that z_t is an i.i.d. process such that

$$z_t \sim \mathcal{N}\left(\mu,\, \sigma_z^2\right). \tag{22}$$

As discussed below, this assumption is correct under FIRE. As in the proposed expectation model, firms are not perfectly aware either of the current value of z_t or of the mean μ of the distribution from which it is drawn. (σ_z^2 is assumed to be known to DM.) Firms' prior beliefs about μ are described as

$$\mu \sim \mathcal{N}(0, \Omega)$$

for some positive Ω.

I denote the average beliefs of firms about z_t and μ as $\hat{z}_t$ and $\hat{\mu}_t$, respectively. Then, $\hat{z}_t$ and $\hat{\mu}_t$ have the following law of motion:

$$\hat{z}_t = \lambda (1 - \kappa) \hat{\mu}_{t-1} + \kappa z_t \tag{23}$$

The average expectation about the mean is

$$\hat{\mu}_t = \lambda (1 - \kappa_\mu) \hat{\mu}_{t-1} + \kappa_\mu z_t. \tag{24}$$

Firms' beliefs are influenced by the realized z_t, which are determined by the rest of the aggregate economy, including the monetary policy discussed in the following section.

7.4 Expectation Formations and Inflation Dynamics

From (20), we see that inflation is determined by the average expectations of firms about the current and future courses of z_t. As seen from (23) and (24), they are are completely specified by two state variables: $\hat{\mu}_{t-1}$ (the average belief about μ in the previous period) and the realized value of z_t. Furthermore, π_t and the exogenous shock e_t determine the evolution of z_t, as described in (21). Combining all these equations, we can deduce that the inflation process is a linear function of e_t and $\hat{\mu}_{t-1}$:

$$\pi_t = \varphi_e e_t + \varphi_\mu \hat{\mu}_{t-1} \tag{25}$$

We can see that π_t is a persistent process since $\hat{\mu}_{t-1}$ is a function of $z_{t-1}, z_{t-2}, \cdots, z_0$, which are in turn functions of lags of π_t and e_t. The coefficients φ_e and φ_μ are derived as

$$\varphi_e = \frac{\delta}{1 + \delta \sigma s + \frac{1}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}}$$

$$\varphi_\mu = \frac{1}{1 + \delta \sigma s + \frac{1}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}} \frac{1-\alpha}{\alpha} \frac{\hat{b}}{\hat{\kappa}},$$

where $\delta \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$, $\hat{\kappa} = \kappa + \kappa_\mu$, and $\hat{b} = \lambda(1-\kappa) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)$. See Appendix G for detailed derivation.

Comparison of different expectation assumptions. Different assumptions about expectation formation result in different inflation dynamics, as captured by $\hat{\kappa}$ and $\hat{b}$. I compare three cases: FIRE ($\phi_n = 0$ and $\phi_m = 0$), the conventional models of information frictions ($\phi_n > 0$ and $\phi_m = 0$), and finally the proposed expectation model ($\phi_n > 0$ and $\phi_m > 0$).

Under FIRE, firms are perfectly aware of z_t and μ. Therefore, firms expect the future marginal costs to be zero on average (since $\mu = 0$) and set their prices to match the current marginal costs. Therefore, the aggregate inflation is proportional to the realized z_t. The inflation process is derived as follows:

$$\pi_t = \frac{\delta}{1 + \delta\,\sigma\,s}\,e_t$$

Under conventional models of information frictions, firms are imperfectly aware of z_t but have come to learn the true mean of the distribution z_t is drawn from. Thus, their subjectively optimal price is equal to the perceived value of the current marginal costs. This is because they correctly expect that their future marginal costs are zero on average. The inflation process is derived as

$$\pi_t = \frac{\delta}{1 + \delta\,\sigma\,s + \frac{1}{\alpha}\,\frac{1-\kappa^*}{\kappa^*}}\,e_t,$$

where κ^* refers to the Kalman gain when updating firms' belief about z_t under the perfect-memory assumption. Firms' reset prices are less responsive to the realized cost-push shocks than under FIRE. This is because firms are not perfectly aware of them when resetting prices. Accordingly, while aggregate inflation is still proportional to the cost-push shocks, the dependence is more muted.

Under the proposed model, the inflation process is derived as follows:

$$\pi_t = \rho_\mu\,\pi_{t-1} + \gamma_0\,e_t + \gamma_1\,e_{t-1}$$

Here, the coefficients on the cost-push shocks are derived as $\gamma_0 = \varphi_e + \varphi_m \kappa_\mu$ and $\gamma_1 = -\varphi_e\,\lambda\,(1 - \kappa_\mu)$. Inflation is persistent, unlike in the previous two expectation models. This is because of the fluctuating beliefs about the long run, as the coefficient ρ_μ is the serial correlation of $\hat{\mu}_t$.

7.5 Calibration

I now discuss how I choose the model parameters. The parameters describing the expectation process come from the previous estimation section. I take the median estimates across macroeconomic variables. For the baseline model, I use $\phi_n = 0.34$, $\phi_m = 0.22$, and $\Omega/\sigma_y^2 = 0.12$. For the conventional models of information frictions, I use $\phi_n = 0.15$.

I set $\chi = 2$ to reflect that the elasticity coefficients of the consumption and labor utility function are both one, following the discussion in Hazell et al. (2022). I assume that firms discount their future revenues with $\beta = 0.99$ because I consider the time unit of the model to be a quarter. The frequency of price changes is matched to the slope of the Phillips curve estimated in the literature.

The inflation response to a 1% increase in the output gap (holding the expectation terms) is estimated to be 0.024 in Rotemberg and Woodford (1997) and 0.0062 in Hazell et al. (2022). I target 0.01 as a midpoint.

Finally, I pin down s in the central bank's targeting rule (19) and the variance of the cost-push shock σ_e^2 to match the empirical volatility of inflation and the output gap. I use the quarterly log changes of the CPI for inflation. For the output gap, I use the difference between the log of real gross domestic product (RGDP) and the log of potential RGDP. All data are from Federal Reserve Economic Data (FRED). The standard deviation of the CPI is 0.35% per quarter, and the standard deviation of the output gap is 2.48% per quarter.

7.6 Monetary Policy and Inflation Variability

We have seen that the expectation-formation process shapes inflation dynamics. In this section, I consider the effects of alternative monetary policies on inflation variability and the role of expectation formation. To do so, I consider values of s in (19) given by

$$s = s^* \frac{\theta}{1-\theta}, \tag{26}$$

where s^* is the calibrated value of s. Thus, $\theta = \frac{1}{2}$ represents the typical monetary policy, bringing the model-predicted volatility of inflation and output closer to the data.

The strength of inflation targeting is measured by $\theta \in [0, 1]$. Complete inflation stabilization is captured by $\theta = 1$. In this case, in response to inflationary pressures from the cost-push shock, the central bank drives output far below the efficient level to stabilize inflation.

The top left panel in Figure 8 shows firms' subjective uncertainty about the long run. The x-axis corresponds to the strength of inflation targeting. I discuss the prediction for three different expectation assumptions: FIRE (black dotted line), the conventional models of information frictions (blue solid line), and the baseline model (orange solid line). As discussed earlier, firms are perfectly aware of the long run under FIRE and the conventional information-frictions model for any monetary-policy rule, but this prediction changes when noisy memory is also present. Firms continually feel uncertain about the long run and keep revising their views. In particular, the strength of inflation targeting matters; more stable inflation means more stable marginal costs, so firms become less uncertain about the long-run mean.

The top right panel of Figure 8 displays the inflation variability for a given monetary-policy rule on the x-axis. I confirm that stronger inflation targeting stabilizes the inflation process for all expectation assumptions. Furthermore, we can see that conventional information-friction models predict more stable inflation than under FIRE. Since firms are not perfectly aware of the realized marginal cost, they do not reflect it in their prices. In the baseline model, firms are imperfectly aware of both the realized marginal cost and its long-run mean. Therefore, their expectations of future marginal costs fluctuate, inducing more price fluctuations.

The bottom panel in Figure 8 illustrates the central bank's trade-off in simultaneously stabilizing inflation and the output gap. Under the conventional information-frictions model, the policy frontier shifts inward compared to FIRE; the economy faces less variable inflation at any output variability. In the baseline model, the policy frontier shifts out, indicating that for any output variability, the economy bears more variable inflation.

7.7 Efficient Inflation Targeting

We have seen that the effect of monetary policy on inflation variability varies with the expectation assumptions. In this section, I study the efficient level of inflation targeting that maximizes social welfare for each expectation assumption.

Let us assume that social welfare depends on how variable the output gap and inflation are. Let us further assume that the welfare-relevant measure of the output gap is the output gap scaled by $\frac{1}{s^*}$. Thus, the welfare losses from the output gap and inflation are roughly comparable in size. Thus:

$$\mathcal{L} = (1 - \omega)\ Var\left[\tilde{x}_t\right] + \omega\, Var\left[\pi_t\right] \tag{27}$$

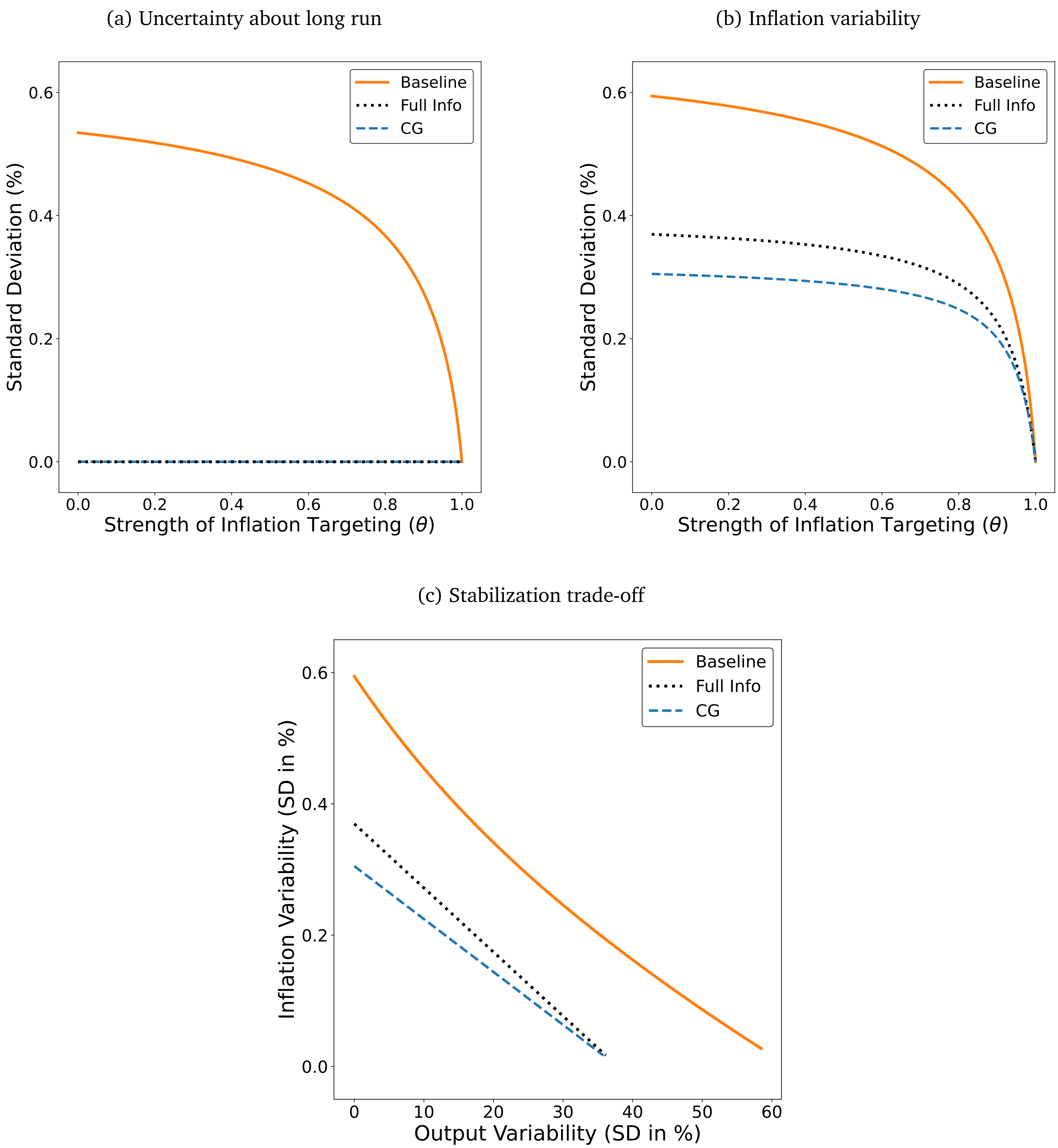
Here, $\tilde{x}_t = \frac{1}{s^*}\, x_t$, and ω reflects the central bank's preference for stabilizing inflation over stabilizing the output gap. I find the optimal level of θ that minimizes the loss function.[10]

The left panel in Figure 9 displays the efficient weight for a given ω (the welfare weight on inflation). Under the conventional information-frictions assumption, it is efficient to put less emphasis on inflation targeting than under FIRE. Since the inflation process is less responsive to fluctuations in marginal cost, the central bank can put more weight on stabilizing the output gap. In comparison, putting more weight on inflation is efficient in the baseline model. Since the volatile inflation process feeds into more widely fluctuating beliefs about the long-run economy, the central bank prioritizes stabilizing inflation.

The right panel in Figure 9 illustrates that conducting monetary policy based on a correct expectation assumption is essential. I show the additional inflation variability that the economy incurs if the central bank adopts a monetary policy that is only efficient under different expectation assumptions. The increased volatility is especially sizable when the central bank intends to produce more stable inflation (that is, when the welfare weight on inflation is high). That is, the central bank can generate volatile inflation because it is not cognizant that fluctuation in marginal costs will unanchor long-run expectations.

[10]One can consider the welfare-loss function whose measure of the output gap is not scaled. I propose to use the scaled output gap to see the effect of different expectation assumptions more clearly for the entire range of ω. In the current exercise, the standard deviation of the output gap is more than seven times larger than that of inflation. Thus, the efficient strength of inflation targeting is quite small unless the welfare weight on inflation is sizable.

Figure 8: The effect of monetary policy

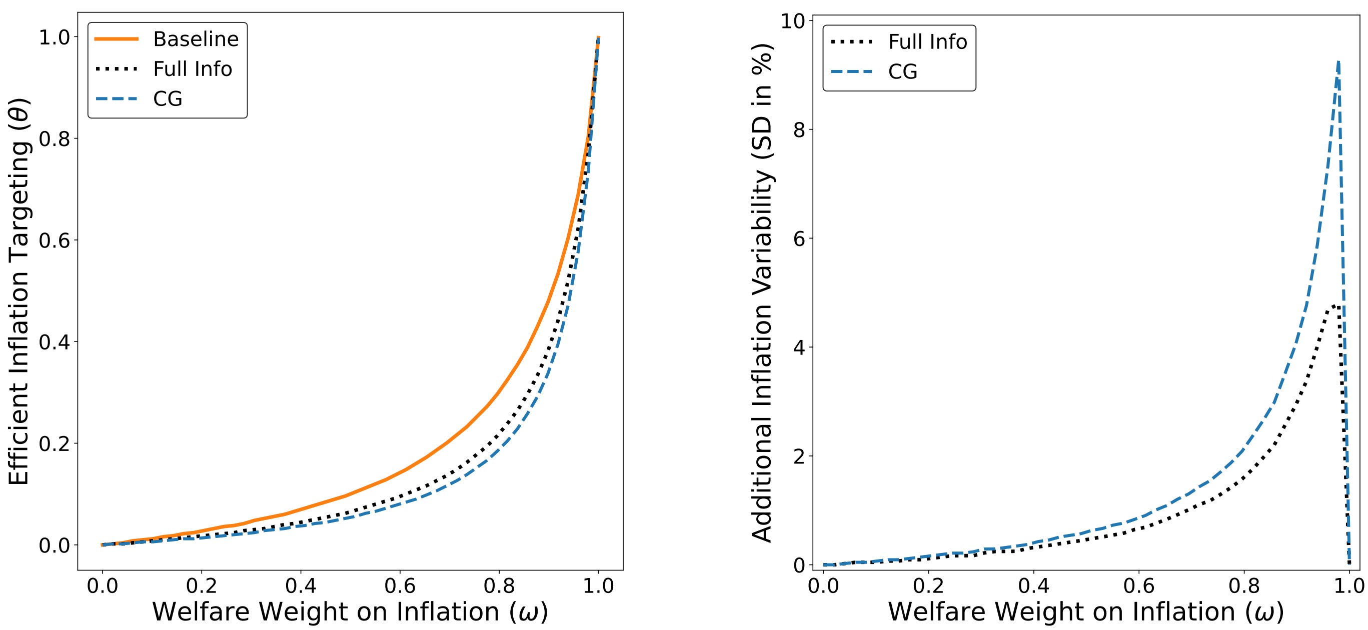


The figures above illustrate the macroeconomic dynamics for varying degrees of strength of inflation targeting (θ). For all figures, three lines correspond to different expectation-formation assumptions: "Baseline" is the proposed model, "Full Info" is the full-information model, and "CG" is the conventional models of information frictions. For each targeting rule θ on the x-axis, the top left panel displays the uncertainty about the long-run mean μ, and the top right panel shows the inflation variability. The bottom panel reports the policy trade-off between inflation stabilization and output-gap stabilization. The model parameters are stated in the main text.

Figure 9: Efficient policy

(a) Efficient inflation targeting

(b) Increased inflation volatility from naive policy



The left panel shows the efficient level of inflation targeting that minimizes the welfare loss (27). The welfare weight on inflation variability (ω) is on the x-axis. Three lines correspond to different expectation-formation assumptions as in Figure 8. The right panel shows the increased inflation variability from implementing inefficient targeting rules. The targeting rules are only efficient if expectations are not subject to noisy memory.

8 Conclusion

I proposed an expectations model in which economic agents make forecasts subject to information frictions. The proposed model accounts for puzzling patterns that conventional information-friction models cannot. It also offers an estimation strategy to identify the extent of information frictions. Using professional forecasters' overall projections of the US economy, I showed that an influential methodology previously proposed in the literature underestimates the extent of information frictions by half. Using the estimated model, I discussed the model's implications for inflation expectations and monetary policy. The public's expectations about the long-run state of the economy are not as well anchored as conventional information-friction models predict. I showed that the central bank's emphasis on inflation stabilization can be more desirable.

To reach these findings, I proposed that the relevant information friction is the cognitive constraint in processing the vast amount of information people have access to. Importantly, I proposed that economic agents process information both external and internal to their minds. This contrasts with conventional information-friction models, which implicitly assume that internal information is perfectly accessible. I showed that jointly considering the two information constraints is crucial to correctly estimating the extent of information frictions. To study the macroeconomic implications, I introduced the proposed expectation model into a standard New Keynesian model. I showed that price-setting firms have unanchored expectations about the long run when internal information is not perfectly accessible. Furthermore, I showed that policies that are efficient under conventional information-friction models generate excessive inflation volatility.

An important lesson from my analysis is that it is crucial to identify the fundamental bottleneck that keeps economic agents from making forecasts consistent with FIRE. I showed that finite capacity to process information — both external and internal — explains various features of survey forecasts that previous expectation-formation models cannot. Recognition of these constraints allows one to see that conventional assumptions in macroeconomic models may not be well grounded. One example is the assumption that agents would be well aware of the long-run economic trends if the economy were stable. In the proposed model, agents' long-run expectations perpetually fluctuate even after extensive learning opportunities. This has the crucial implication that seemingly anchored long-run inflation expectations can start moving when agents witness bouts of high inflation. Thus, a monetary authority whose policies rely on the prospect of firmly anchored expectations can lose its grip on the economy, leaving economic agents to doubt the authority's ability to manage inflation. Empirically relevant expectation-formation models can guide the complex considerations that conducting monetary policy requires, especially in new environments yet to be experienced and analyzed.

9 Accompanying Tables and Figures

Table 3: Estimated regression coefficients using OLS

	β_C	SE	p-value	β_I	SE	p-value	β_K	SE	p-value
Nominal Gross Domestic Product	0.63	0.11	0.0	-0.27	0.04	0.0	0.54	0.03	0.0
Real Gross Domestic Product	0.45	0.12	0.0	-0.25	0.05	0.0	0.6	0.02	0.0
GDP Chain-Weighted Price Index	0.71	0.11	0.0	-0.32	0.04	0.0	0.6	0.03	0.0
Corporate Profits after Taxes	0.68	0.17	0.0	-0.44	0.05	0.0	0.51	0.03	0.0
Civilian Unemployment Rate	0.62	0.08	0.0	-0.05	0.05	0.31	0.62	0.02	0.0
Industrial Production Index	0.61	0.13	0.0	-0.18	0.06	0.0	0.58	0.02	0.0
Housing Starts	0.5	0.12	0.0	-0.25	0.06	0.0	0.58	0.02	0.0
Consumer Price Index	0.55	0.15	0.0	-0.17	0.09	0.04	0.56	0.03	0.0
Treasury Bill Rate, 3-month	0.29	0.05	0.0	-0.01	0.04	0.85	0.73	0.03	0.0
AAA Corporate Bond Yield	0.05	0.07	0.48	-0.35	0.04	0.0	0.68	0.02	0.0
Treasury Bond Rate, 10-year	0.28	0.07	0.0	-0.12	0.05	0.01	0.7	0.03	0.0

The first column shows the variables included in the SPF's "U.S. Business Indicators" section. The first panel displays the estimated regression coefficient from (13). The standard errors are robust to the presence of arbitrary heteroskedasticity and autocorrelation. The second panel shows the regression coefficient estimates from (14) when individual forecasts are pooled. For this regression, the standard errors are two-way clustered by forecasters and survey date. The last panel reports the regression coefficient from (15). The standard errors are similarly clustered two-way. I include individual and horizon fixed effects for the last two regression specifications.

Table 4: Estimated parameters

	ϕ_n	CI	ϕ_m	CI	Ω/σ_y^2	CI	σ_ν^2/σ_y^2	CI
Nominal Gross Domestic Product	0.34	(0.29,0.39)	0.19	(0.15,0.24)	0.16	(0.14,0.18)	0.14	(0.09,0.22)
Real Gross Domestic Product	0.32	(0.28,0.36)	0.27	(0.21,0.34)	0.12	(0.11,0.14)	0.23	(0.15,0.35)
GDP Chain-Weighted Price Index	0.53	(0.46,0.59)	0.45	(0.36,0.52)	0.1	(0.09,0.11)	0.52	(0.25,1.0)
Corporate Profits after Taxes	0.51	(0.43,0.56)	0.44	(0.41,0.48)	0.08	(0.08,0.09)	1.0	(1.0,1.0)
Civilian Unemployment Rate	0.14	(0.13,0.2)	0.0	(0.0,0.1)	1.0	(0.14,1.0)	0.0	(0.0,0.03)
Industrial Production Index	0.27	(0.19,0.32)	0.14	(0.03,0.2)	0.19	(0.14,1.0)	0.1	(0.07,0.16)
Housing Starts	0.38	(0.32,0.47)	0.3	(0.21,0.4)	0.12	(0.1,0.14)	0.32	(0.18,0.64)
Consumer Price Index	0.34	(0.27,0.4)	0.22	(0.1,0.33)	0.12	(0.09,0.27)	0.28	(0.1,0.58)
Treasury Bill Rate, 3-month	0.11	(0.06,0.17)	0.0	(0.0,0.0)	0.27	(0.16,1.0)	0.09	(0.03,0.39)
AAA Corporate Bond Yield	0.38	(0.31,0.42)	0.55	(0.48,0.6)	0.08	(0.08,0.09)	1.0	(0.61,1.0)
Treasury Bond Rate, 10-year	0.14	(0.12,0.19)	0.09	(0.03,0.21)	0.26	(0.12,0.92)	0.08	(0.03,0.13)

Each panel shows the estimated parameter and its confidence interval. The standard error is computed by panel-bootstrapping the SPF individual-forecast data. I report the 5% and the 95% point estimates from the bootstrapped samples. I estimate the scaled value of σ_ν^2 and Ω (divided by the variance of the forecast variable). I restrict these scaled σ_ν^2 and Ω to be between zero and one.

Table 5: Estimated parameters using Coibion and Gorodnichenko (2015) approach

	ϕ_n	CI	σ_ν^2/σ_y^2	CI
Nominal Gross Domestic Product	0.17	(0.15,0.19)	0.05	(0.02,0.08)
Real Gross Domestic Product	0.15	(0.12,0.17)	0.07	(0.04,0.12)
GDP Chain-Weighted Price Index	0.21	(0.18,0.24)	0.02	(0.0,0.06)
Corporate Profits after Taxes	0.26	(0.2,0.31)	0.17	(0.08,0.27)
Civilian Unemployment Rate	0.21	(0.19,0.23)	0.01	(0.0,0.03)
Industrial Production Index	0.19	(0.17,0.21)	0.04	(0.02,0.08)
Housing Starts	0.17	(0.14,0.2)	0.09	(0.05,0.13)
Consumer Price Index	0.23	(0.18,0.28)	0.11	(0.04,0.19)
Treasury Bill Rate, 3-month	0.11	(0.08,0.12)	0.03	(0.0,0.06)
AAA Corporate Bond Yield	0.02	(0.0,0.05)	0.16	(0.13,0.2)
Treasury Bond Rate, 10-year	0.11	(0.09,0.13)	0.05	(0.02,0.08)

Each panel shows the estimated parameter and its confidence interval. The standard error is computed by panel-bootstrapping the SPF individual-forecast data. I report the 5% and the 95% point estimates from the bootstrapped samples. I estimate the scaled value of σ_ν^2 (divided by the variance of the forecast variable). I restrict these scaled σ_ν^2 to be between zero and one.

Table 6: Model fit

(a) Targeted Moments

	β_C		β_I		β_K		$\beta_{\mu,K}$	
	Data	Model	Data	Model	Data	Model	Data	Model
Nominal Gross Domestic Product	0.55	0.54	-0.27	-0.27	0.55	0.55	0.08	0.08
Real Gross Domestic Product	0.36	0.35	-0.24	-0.24	0.61	0.61	0.08	0.08
GDP Chain-Weighted Price Index	0.56	0.57	-0.32	-0.32	0.6	0.6	0.08	0.08
Corporate Profits after Taxes	0.49	0.5	-0.44	-0.27	0.51	0.57	0.08	0.08
Civilian Unemployment Rate	0.56	0.59	-0.05	-0.05	0.63	0.65	0.08	0.04
Industrial Production Index	0.53	0.52	-0.18	-0.18	0.57	0.58	0.08	0.08
Housing Starts	0.41	0.4	-0.27	-0.27	0.58	0.58	0.08	0.08
Consumer Price Index	0.46	0.47	-0.17	-0.17	0.57	0.57	0.08	0.08
Treasury Bill Rate, 3-month	0.28	0.28	-0.01	-0.01	0.73	0.63	0.08	0.03
AAA Corporate Bond Yield	0.03	0.02	-0.35	-0.34	0.68	0.68	0.08	0.08
Treasury Bond Rate, 10-year	0.26	0.25	-0.12	-0.11	0.7	0.7	0.08	0.08

(b) Not-targeted moments

	Variation in Forecasts				Variation in Revisions			
	Time Series		Cross Section		Time Series		Cross Section	
	Data	Model	Data	Model	Data	Model	Data	Model
Nominal Gross Domestic Product	0.73	0.68	0.4	0.41	0.22	0.26	0.39	0.43
Real Gross Domestic Product	0.59	0.63	0.39	0.39	0.26	0.33	0.4	0.42
GDP Chain-Weighted Price Index	0.78	0.4	0.31	0.41	0.17	0.25	0.3	0.44
Corporate Profits after Taxes	0.62	0.35	0.74	0.35	0.31	0.26	0.73	0.36
Civilian Unemployment Rate	0.51	0.75	0.3	0.33	0.23	0.34	0.31	0.37
Industrial Production Index	0.57	0.71	0.4	0.37	0.27	0.31	0.4	0.39
Housing Starts	0.71	0.57	0.55	0.4	0.38	0.31	0.55	0.42
Consumer Price Index	0.38	0.56	0.21	0.37	0.16	0.34	0.2	0.38
Treasury Bill Rate, 3-month	0.54	0.78	0.29	0.25	0.39	0.39	0.34	0.27
AAA Corporate Bond Yield	0.72	0.44	0.73	0.35	0.58	0.34	0.76	0.4
Treasury Bond Rate, 10-year	0.49	0.78	0.43	0.3	0.43	0.39	0.47	0.34

The table compares the predictions of the estimated model to the data moments. The upper panel shows the targeted moments, and the lower panel shows untargeted moments. For untargeted moments, I report variations of forecasts and forecast revisions in the time series and cross sections, whose units are standard deviations scaled by the standard deviation of the forecast variables. For each macroeconomic variable, these moments are averaged across four consecutive forecast horizons (current to three quarters ahead).

References

ADAM, K., A. MARCET, AND J. BEUTEL (2017): "Stock Price Booms and Expected Capital Gains," *American Economic Review*, 107, 2352–2408.

AFROUZI, H., S. Y. KWON, A. LANDIER, Y. MA, AND D. THESMAR (2020): "Overreaction in Expectations: Evidence and Theory," SSRN Working Paper.

AFROUZI, H. AND C. YANG (2021): "Dynamic Rational Inattention and the Phillips Curve," Working Paper.

ANDRADE, P., R. K. CRUMP, S. EUSEPI, AND E. MOENCH (2016): "Fundamental disagreement," *Journal of Monetary Economics*, 83, 106–128.

ANDRADE, P. AND H. LE BIHAN (2013): "Inattentive professional forecasters," *Journal of Monetary Economics*, 60, 967–982.

ANGELETOS, G.-M., F. COLLARD, AND H. DELLAS (2020): "Business-Cycle Anatomy," *American Economic Review*, 110, 3030–3070.

ANGELETOS, G.-M. AND Z. HUO (2021): "Myopia and Anchoring," *American Economic Review*, 111, 1166–1200.

ANGELETOS, G.-M., Z. HUO, AND K. A. SASTRY (2021): "Imperfect Macroeconomic Expectations: Evidence and Theory," *NBER Macroeconomics Annual*, 35, 1–86.

ANGELETOS, G.-M. AND J. LA'O (2013): "Sentiments," *Econometrica*, 81, 739–779.

ANGELETOS, G.-M. AND C. LIAN (2018): "Forward Guidance without Common Knowledge," *The American Economic Review*, 108, 2477–2512.

AZEREDO DA SILVEIRA, R., Y. SUNG, AND M. WOODFORD (2020): "Optimally Imprecise Memory and Biased Forecasts," NBER Working Paper No. 28075.

BAKER, S., T. MCELROY, AND X. SHENG (2020): "Expectation Formation Following Large, Unexpected Shocks," *The Review of Economics and Statistics*, 102, 287–303.

BARBERIS, N., R. GREENWOOD, L. JIN, AND A. SHLEIFER (2015): "X-CAPM: An extrapolative capital asset pricing model," *Journal of Financial Economics*, 115, 1–24.

BIANCHI, F., S. C. LUDVIGSON, AND S. MA (2022): "Belief Distortions and Macroeconomic Fluctuations," Working Paper.

BORDALO, P., J. J. CONLON, N. GENNAIOLI, S. Y. KWON, AND A. SHLEIFER (2022): "emory and Probability," *The Quarterly Journal of Economics*.

BORDALO, P., N. GENNAIOLI, R. LA PORTA, AND A. SHLEIFER (2020a): "Expectations of Fundamentals and Stock Market Puzzles," NBER Working Paper No. 27283.

BORDALO, P., N. GENNAIOLI, Y. MA, AND A. SHLEIFER (2020b): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110, 2748–2782.

BORDALO, P., N. GENNAIOLI, R. L. PORTA, AND A. SHLEIFER (2019): "Diagnostic Expectations and Stock Returns," *The Journal of Finance*, 74, 2839–2874.

BOUCHAUD, J.-P., P. KRÜGER, A. LANDIER, AND D. THESMAR (2019): "Sticky Expectations and the Profitability Anomaly," *The Journal of Finance*, 74, 639–674.

BROER, T. AND A. N. KOHLHAS (2021): "Forecaster (Mis-)Behavior," Working Paper.

CARVALHO, C., S. EUSEPI, E. MOENCH, AND B. PRESTON (2022): "Anchored Inflation Expectations," Working Paper.

CASE, K. E., R. J. SHILLER, AND A. THOMPSON (2012): "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets," NBER Working Paper No. 18400.

CIESLAK, A. (2018): "Short-Rate Expectations and Unexpected Returns in Treasury Bonds," *The Review of Financial Studies*, 31, 3265–3306.

COIBION, O., U. AUSTIN, Y. GORODNICHENKO, AND M. WEBER (2021): "MONETARY POLICY COMMUNICATIONS AND THEIR EFFECTS ON HOUSEHOLD INFLATION EXPECTATIONS," NBER Working Paper No. 25482.

COIBION, O. AND Y. GORODNICHENKO (2012): "What Can Survey Forecasts Tell Us about Information Rigidities?" *Journal of Political Economy*, 120, 116–159.

——— (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *The American Economic Review*, 105, 2644–2678.

COIBION, O., Y. GORODNICHENKO, AND S. KUMAR (2018): "How Do Firms Form Their Expectations? New Survey Evidence," *The American Economic Review*, 108, 2671–2713.

COLLIN-DUFRESNE, P., M. JOHANNES, AND L. A. LOCHSTOER (2016): "Parameter Learning in General Equilibrium: The Asset Pricing Implications," *American Economic Review*, 106, 664–698.

D'ARIENZO, D. (2020): "Maturity Increasing Overreaction and Bond Market Puzzles," SSRN Working Paper.

ELLIOTT, G., I. KOMUNJER, AND A. TIMMERMANN (2008): "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association*, 6, 122–157.

ENKE, B. AND T. GRAEBER (2019): "Cognitive Uncertainty," NBER Working Paper No. 26518.

EUSEPI, S. AND B. PRESTON (2011): "Expectations, Learning, and Business Cycle Fluctuations," *American Economic Review*, 101, 2844–2872.

FARMER, L., E. NAKAMURA, AND J. STEINSSON (2021): "Learning About the Long Run," .

FUSTER, A., D. LAIBSON, AND B. MENDEL (2010): "Natural Expectations and Macroeconomic Fluctuations," *The Journal of Economic Perspectives*, 24, 67–84.

GALÍ, J. (2008): *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework*, Princeton University Press.

GÀTI, L. (2021): "Monetary Policy & Anchored Expectations An Endogenous Gain Learning Model," Working Paper.

GEMMI, L. AND R. VALCHEV (2021): "Biased Surveys," Working Paper.

GIGLIO, S. AND B. KELLY (2018): "Excess Volatility: Beyond Discount Rates," *The Quarterly Journal of Economics*, 133, 71–127.

GREENWOOD, R. AND A. SHLEIFER (2014): "Expectations of Returns and Expected Returns," *Review of Financial Studies*, 27, 714–746.

HAZELL, J., J. HERREÑO, E. NAKAMURA, AND J. STEINSSON (2022): "The Slope of the Phillips Curve: Evidence from U.S. States," *The Quarterly Journal of Economics*, 137.

JUODIS, A. AND S. KUCINSKAS (2019): "Quantifying Noise," SSRN Working Paper.

K. CRUMP, R., S. EUSEPI, E. MOENCH, AND B. PRESTON (2021): "The Term Structure of Expectations," Federal Reserve Bank of New York Staff Reports.

KOHLHAS, A. N. AND A. WALTHER (2021): "Asymmetric Attention," *American Economic Review*, 111, 2879–2925.

MACKOWIAK, B. AND M. WIEDERHOLT (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99, 769–803.

MANKIW, N. G. AND R. REIS (2002): "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *The Quarterly Journal of Economics*, 117, 1295–1328.

MANSKI, C. F. (2018): "Survey Measurement of Probabilistic Macroeconomic Expectations: Progress and Promise," *NBER Macroeconomics Annual*, 32, 411–471.

MCELROY, T. AND X. S. SHENG (2021): "Augmented Information Rigidity Test," SSRN Working Paper.

MIAO, J., J. WU, AND E. R. YOUNG (2022): "Multivariate Rational Inattention," *Econometrica*, 90, 907–945.

NAGEL, S. AND Z. XU (2021): "Asset Pricing with Fading Memory," *The Review of Financial Studies*, 00, 56.

NORDHAUS, W. D. (1987): "Forecasting Efficiency: Concepts and Applications," *The Review of Economics and Statistics*, 69, 667–674.

ORPHANIDES, A. AND J. WILLIAMS (2004): *Imperfect Knowledge, Inflation Expectations, and Monetary Policy*, University of Chicago Press.

ORPHANIDES, A. AND J. C. WILLIAMS (2006): "Monetary Policy with Imperfect Knowledge," *Journal of the European Economic Association*, 4, 366–375.

PATTON, A. J. AND A. TIMMERMANN (2010): "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion," *Journal of Monetary Economics*, 57, 803–820.

PESARAN, M. H. AND M. WEALE (2006): "Chapter 14 Survey Expectations," in *Handbook of Economic Forecasting*, Elsevier, vol. 1, 715–776.

PIAZZESI, M. AND M. SCHNEIDER (2009): "Trend and Cycle in Bond Premia," Working Paper.

REIS, R. (2020): "The People versus the Markets: A Parsimonious Model of Inflation Expectations," CEPR Discussion Papers No. 15624.

ROTEMBERG, J. AND M. WOODFORD (1997): *An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy*, National Bureau of Economic Research, Inc, 297–361.

SIMS, C. A. (2003): "Implications of rational inattention," *Journal of Monetary Economics*, 50, 665–690.

SINGLETON, K. J. (2021): "Presidential Address: How Much "Rationality" Is There in Bond-Market Risk Premiums?" *The Journal of Finance*, 76, 1611–1654.

TIMMERMANN, A. G. (1993): "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *The Quarterly Journal of Economics*, 108, 1135–1145.

WACHTER, J. A. AND M. J. KAHANA (2021): "A retrieved-context theory of financial decisions," SSRN Working Paper.

WACHTER, J. A. AND U. MALMENDIER (2022): "Memory of Past Experiences and Economic Decisions," Working Paper.

WANG, C. (2021): "Under- and Overreaction in Yield Curve Expectations," SSRN Working Paper.

WEITZMAN, M. L. (2007): "Subjective Expectations and Asset-Return Puzzles," *American Economic Review*, 97, 1102–1130.

WOODFORD, M. (2003): "Imperfect common knowledge and the effects of monetary policy," *Knowledge, information, and expectations in modern macroeconomics : in honor of Edmund S. Phelps*, 25–58, Princeton University Press.

——— (2020): "Modeling Imprecision in Perception, Valuation, and Choice," *Annual Review of Economics*, 12, 579–601.

ZARNOWITZ, V. (1985): "Rational Expectations and Macroeconomic Forecasts," *Journal of Business & Economic Statistics*, 3, 293–311.

A Derivation of the Optimal Cognitive Process

For any given state vector x_t, I show the optimal structure of the cognitive process, described by the sequence of $\{K_t, \sigma_{u,t}, \Lambda_t, \sigma_{\omega,t}\}_{t=0}^{\infty}$, that minimizes the loss function (7) subject to the information environment (3), (4), (5), and (6).

A.1 Proof: The Optimal Structure for the Representation

I show below that the optimal $n_{i,t}$ and $m_{i,t}$ are one-dimensional. In particular, I show that the optimal $n_{i,t}$ records $E\left[z_t | N_t\right]$ with noise while the optimal $m_{i,t}$ stores $z_{i,t|t-1}$ with noise.

Step 1: Partition of $n_{i,t}$ and $m_{i,t}$

Partition of $n_{i,t}$ We can partition $n_{i,t} = K_t \cdot N_t + u_{i,t}$ into the following form

$$\begin{pmatrix} \vec{n}_{i,t} \\ \tilde{n}_{i,t} \end{pmatrix} = \begin{pmatrix} K_{a,t} & K_{b,t} \\ K_{c,t} & K_{d,t} \end{pmatrix} \begin{pmatrix} \vec{N}_t \\ E\left[x | N_t\right] \end{pmatrix} + \begin{pmatrix} \vec{u}_{i,t+1} \\ \tilde{u}_{i,t+1} \end{pmatrix} \tag{28}$$

Note that the elements of $\vec{N}_t$ are not correlated with $E\left[x_t | N_t\right]$ and that $\vec{N}_t$ and $E\left[x_t | N_t\right]$ span the same vector space as N_t. I also impose the following normalization assumption

$$E\left[x_t | m_{i,t}, n_{i,t}\right] = \tilde{n}_{i,t} + cons \cdot E\left[x_t | m_{i,t}\right]$$

This relationship holds if and only if $E\left[x_t | N_t\right] - \tilde{n}_{i,t}$ is uncorrelated with all the elements in $n_{i,t}$ conditional on $m_{i,t}$. That is, the two requirements are

$$Cov\left[x_t - \tilde{n}_{i,t}, \vec{n}_{i,t} | m_{i,t}\right] = \vec{O} \tag{29a}$$

$$Cov\left[x_t - \tilde{n}_{i,t}, \tilde{n}_{i,t} | m_{i,t}\right] = O \tag{29b}$$

We can see that (29b) implies

$$\begin{aligned} & Cov\left[x_t, K_{d,t}\, E\left[x_t | N_t\right] | m_{i,t}\right] = Var\left[K_{c,t}\, \vec{N}_t + K_{d,t}\, E\left[x_t | N_t\right] + \tilde{u}_{i,t} \middle| m_{i,t}\right] \\ \Leftrightarrow \quad & Var\left[K_{c,t}\, \vec{N}_t + \tilde{u}_{i,t} \middle| m_{i,t}\right] = Cov\left[x_t, K_{d,t}\, E\left[x_t | N_t\right] | m_{i,t}\right] - K_{d,t}\, Var\left[E\left[x_t | N_t\right] | m_{i,t}\right] K_{d,t}' \end{aligned}$$

The feasible set of $K_{d,t}$ is defined as $K_{d,t}$ that yields the right-hand-side term to be a proper variance-covariance matrix (that is, symmetric and p.s.d.).

Partition of $m_{i,t}$ Similarly, we can also partition $m_{i,t} = \Lambda_t \cdot \begin{pmatrix} m_{i,t-1} \\ n_{i,t-1} \end{pmatrix} + \omega_{i,t}$ as the following form

$$\begin{pmatrix} \vec{m}_{i,t} \\ \tilde{m}_{i,t} \end{pmatrix} = \begin{pmatrix} \Lambda_{a,t} & \Lambda_{b,t} \\ \Lambda_{c,t} & \Lambda_{d,t} \end{pmatrix} \begin{pmatrix} \vec{s}_{i,t-1} \\ x_{i,t|t-1} \end{pmatrix} + \begin{pmatrix} \vec{\omega}_{i,t} \\ \tilde{\omega}_{i,t} \end{pmatrix} \tag{30}$$

Note that the elements of $\vec{s}_{i,t}$ are not correlated with $z_{i,t|t-1}$ and that $\vec{s}_{i,t-1}$ and $x_{i,t|t-1}$ span the same vector space as $(m_{i,t-1}, n_{i,t-1})$. I also impose the following normalization assumption

$$E\left[x_{i,t|t-1}\middle| m_{i,t}\right] = \tilde{m}_{i,t} + cons \cdot E\left[x_{i,t|t-1}\right]$$

This relationship holds if and only if $x_{i,t|t-1} - \tilde{m}_{i,t}$ is uncorrelated with all the elements in $m_{i,t}$. Two requirements summarize this relationship.

$$Cov\left[x_{i,t|t-1} - \tilde{m}_{i,t}, \vec{m}_{i,t}\right] = \vec{O} \tag{31a}$$

$$Cov\left[x_{i,t|t-1} - \tilde{m}_{i,t}, \tilde{m}_{i,t}\right] = O \tag{31b}$$

The second requirement implies that

$$\begin{aligned} & Cov\left[x_{i,t|t-1}, \tilde{m}_{i,t}\right] = Var\left[\tilde{m}_{i,t}\right] \\ \Leftrightarrow \quad & Var\left[\Lambda_{c,t}\, \vec{s}_{i,t-1} + \tilde{\omega}_{i,t}\right] = (1 - \Lambda_{d,t})\, Var\left[x_{i,t|t-1}\right]\, \Lambda'_{d,t} \\ & \qquad = (1 - \Lambda_{d,t}) \left(Var\left[x_t\right] - \Sigma_{t|t-1}\right) \Lambda'_{d,t} \end{aligned}$$

The feasible set of $\Lambda_{d,t}$ is defined as the collection of $\Lambda_{d,t}$ under which the resulting right-hand side is a proper variance-covariance matrix (that is, symmetric and p.s.d.).

Step 2: Forecast accuracy depends only on $K_{d,t}$ and $\Lambda_{d,t}$

From the proposed partition (28), we can see that

$$x_t|\, m_{i,t}, n_{i,t} = x_t|\, m_{i,t}, \tilde{n}_{i,t}$$

That is, further knowledge of $\vec{n}_{i,t}$ does not improve the estimate of $x_t|\, m_{i,t}, \tilde{n}_{i,t}$. This follows from (29a). Furthermore, we can see that $K_{d,t}$ uniquely determines the posterior uncertainty $\Sigma_{t|t}$, given the prior uncertainty $\Sigma^m_{t|t}$,

Likewise, we can also see from (30) that

$$x_{i,t|t-1}\big|\, m_{i,t} = x_{i,t|t-1}\big|\, \tilde{m}_{i,t}$$

The information in $m_{i,t}$ about $x_{i,t|t-1}$ is completely captured by $\tilde{m}_{i,t}$, which follows from (31a). We can furthermore see that $\Lambda_{d,t}$ uniquely determines the next-period prior uncertainty given $\Sigma_{t|t-1}$, the time-t posterior uncertainty about x_t.

$$\begin{aligned} Var\left[x_t \middle| \tilde{m}_t\right] &= Var\left[x_t\right] - \Lambda_{d,t}\, Cov\left[x_t, x_{i,t|t-1}\right] \\ &= Var\left[x_t\right] - \Lambda_{d,t}\left(Var\left[x_t\right] - \Sigma_{x|t-1}\right) \\ &= \left(1 - \Lambda_{d,t}\right) Var\left[x_t\right] + \Lambda_{d,t}\, \Sigma_{t|t-1} \end{aligned}$$

In summary, given $\Sigma_{t|t-1}$ at any time t, $K_{d,t}$ and $\Lambda_{d,t}$ uniquely determine $\Sigma^m_{t|t}$ and $\Sigma_{t|t}$. We can apply this argument recursively. It must be that the sequence of $\{K_{d,t}, \Lambda_{d,t}\}$ uniquely determines the sequence of $\left\{\Sigma_{t|t}\right\}$, given the initial prior uncertainty.

Step 3: The Optimal Choice of K_t and Λ_t

Since the remaining elements of K_t and Λ_t do not matter for the forecast accuracy, we can furthermore conclude that it is optimal to have them equal to zero. To see why note that

$$I\left(n_{i,t}; N_t\right) = I\left(\left(\vec{n}_{i,t}, \tilde{n}_{i,t}\right); \left(\vec{N}_t, E\left[x_t \middle| N_t\right]\right)\right)$$

As discussed in Appendix C.2 of Azeredo da Silveira et al. (2020), the lower bound of this mutual information is equal to $I\left(\tilde{n}_{i,t}; E\left[x_t \middle| N_t\right]\right)$. This lower bound is achieved when $K_{a,t} = K_{b,t} = K_{c,t} = O$. Likewise,

$$I\left(m_{i,t}; m_{i,t-1}, n_{i,t-1}\right) = I\left(\left(\vec{m}_{i,t}, \tilde{m}_{i,t}\right); \left(\vec{s}_{i,t-1}, x_{i,t|t-1}\right)\right)$$

whose lower bound is equal to $I\left(\tilde{m}_{i,t}; x_{i,t|t-1}\right)$. This lower bound is achieved when $\Lambda_{a,t} = \Lambda_{b,t} = \Lambda_{c,t} = O$.

B Optimal Covnitive Process When z_t is the Only State Variable

In this section, I apply the result from the previous section when $x_t = z_t$.

B.1 Optimal representation of noisy news

The optimal $n_{i,t}$ is described as

$$\tilde{n}_{i,t} = \kappa_t \cdot E\left[z_t \middle| N_t\right] + \tilde{u}_{i,t}$$

for some positive scalar κ_t. The idiosyncratic noise $\tilde{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(0,\, \sigma^2_{u,t}\right)$, where $\sigma^2_{u,t}$ is determined by the choice of κ_t.

$$\sigma^2_{u,t} = \kappa_t\, Cov\left[z_t, E\left[z_t \middle| N_t\right] \middle| m_{i,t}\right] - \kappa_t^2\, Var\left[E\left[z_t \middle| N_t\right] \middle| m_{i,t}\right]$$

Without loss of generality, we could assume that $E\left[z_t | N_t\right]$ can be expressed as

$$E\left[z_t | N_t\right] = z_t + \bar{\nu}_t$$

where $\bar{\nu}_t \sim \mathcal{N}\left(0,\, \sigma_\nu^2\right)$ for some positive σ_ν^2. Then, $\sigma_{u,t}^2$ is further simplified to

$$\sigma_{u,t}^2 = \kappa_t \left(1 - \kappa_t\right) \Sigma_{z,t|t}^m - \kappa_t^2\, \sigma_\nu^2$$

where $\Sigma_{z,t|t}^m = Var\left[z_t | m_{i,t}\right]$. Any $\kappa_t \in \left[0, \frac{\Sigma_{z,t|t}^m}{\Sigma_{z,t|t}^m + \sigma_\nu^2}\right]$ ensures that the resulting $\sigma_{u,t}^2$ is non-negative.

Determination of κ_t Using the information constraint, we can derive that

$$\begin{aligned} I\left(n_{i,t}; N_t\right) &= I\left(\tilde{n}_{i,t}; E\left[z_t | N_t\right]\right) \\ &= -\frac{1}{2}\log\left(1 - \frac{\kappa_t^2\, Var\left[z_t + \tilde{\nu}_t\right]}{\kappa_t^2\, Var\left[, z_t + \tilde{\nu}_t\right] + \sigma_{u,t}^2}\right) \\ &= -\frac{1}{2}\log\left(1 - \frac{Var\left[z_t + \tilde{\nu}_t\right]}{Var\left[z_t + \tilde{\nu}_t\right] + \left(\left(\kappa_t^{-1} - 1\right)\, \Sigma_{t|t}^m - \sigma_\nu^2\right)}\right) \le -\frac{1}{2}\log\, \phi_n \end{aligned}$$

Rearranging the last inequality yields

$$\kappa_t \le \frac{\Sigma_{t|t}^m}{\Sigma_{t|t}^m + \frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma_\nu^2\right) + \sigma_\nu^2} \tag{32}$$

The upper bound is the optimal κ_t. Then, the resulting $\sigma_{u,t}^2$ is

$$\sigma_{u,t}^2 = \frac{\left(\Sigma_{t|t}^m\right)^2 \left(\frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma_\nu^2\right)\right)}{\left(\Sigma_{t|t}^m + \frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma_\nu^2\right) + \sigma_\nu^2\right)^2} \tag{33}$$

B.2 Optimal representation of noisy memory

Likewise, we can express the optimal $m_{i,t}$ as

$$\tilde{m}_{i,t} = \lambda_t \cdot z_{i,t|t-1} + \tilde{\omega}_{i,t}$$

for some positive scalar λ_t. The idiosyncratic noise $\tilde{\omega}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(0,\, \sigma_{\omega,t}^2\right)$, whose variance is determined by the choice of λ_t as follows.

$$\sigma_{\omega,t}^2 = \lambda_t \left(1 - \lambda_t\right) Var\left[z_{i,t|t-1}\right]$$

Any $\lambda_t \in [0, 1]$ ensures that the resulting $\sigma_{\omega,t}^2$ is non-negative.

Determination of λ_t Using the information constraint, we can derive that

$$\begin{aligned} I\left(m_{i,t}; m_{i,t-1}, n_{i,t-1}\right) &= I\left(\tilde{m}_{i,t}; z_{i,t|t-1}\right) \\ &= -\frac{1}{2}\log\det\left(1-\lambda_t\right) \leq -\frac{1}{2}\log\,\phi_m \end{aligned}$$

Therefore,

$$\lambda_t \leq 1 - \phi_m$$

The optimal $\lambda_t = 1 - \phi_m$ and the resulting $\sigma^2_{\omega,t} = \phi_m\left(1-\phi_m\right) Var\left[z_{i,t|t-1}\right]$.

C Optimal Cognitive Process When (μ, z_t) is the State Vector

In this section, I apply the result from Section A.1 when $x_t = (\mu, z_t)$.

C.1 Optimal representation of noisy news

The optimal $n_{i,t}$ is described as

$$\tilde{n}_{i,t} = \tilde{K}_t \cdot E\left[x_t | N_t\right] + \tilde{u}_{i,t}$$

for some matrix scalar $\tilde{K}_t$. The idiosyncratic noise $\tilde{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(O,\, \sigma_{u,t}\right)$, where $\sigma_{u,t}$ is determined by the choice of $\tilde{K}_t$.

$$\sigma_{u,t} = Cov\left[x_t, E\left[x_t | N_t\right] | m_{i,t}\right] \tilde{K}'_t - \tilde{K}_t Var\left[E\left[x_t | N_t\right] | m_{i,t}\right] \tilde{K}'_t$$

Note that $E\left[x_t | N_t\right]$ is spanned by $E\left[z_t | N_t\right]$. This is because the news vector N_t is informative about μ only through the information about z_t. Therefore, without loss of generality, we can express $\tilde{n}_{i,t}$ as

$$\tilde{n}_{i,t} = \begin{pmatrix} \frac{\kappa_{\mu,t}}{\kappa_t} \\ 1 \end{pmatrix} \cdot \left(\kappa_t\, E\left[z_t | N_t\right] + \bar{u}_{i,t}\right)$$

where the idiosyncratic noise $\bar{u}_{i,t}$ is drawn from $\mathcal{N}\left(0,\, \sigma^2_{u,t}\right)$. The noisy news structure is then described by three univariate variables, $\kappa_{\mu,t}$, κ_t, and $\sigma^2_{u,t}$, which remain to be specified.

We could furthermore see that the normalization assumption $Cov\left[x_t, \tilde{n}_{i,t} | m_{i,t}\right] = Var\left[\tilde{n}_{i,t} | m_{i,t}\right]$ implies that

$$\kappa_t \begin{pmatrix} \Sigma^m_{\mu,t|t}\left(\kappa_{\mu,t}/\kappa_t\right) & \Sigma^m_{\mu,t|t} \\ \Sigma^m_{z,t|t}\left(\kappa_{\mu,t}/\kappa_t\right) & \Sigma^m_{z,t|t} \end{pmatrix} = \left(\kappa_t^2\left(\Sigma^m_{z,t|t} + \sigma^2_\nu\right) + \sigma^2_{u,t}\right) \begin{pmatrix} \left(\kappa_{\mu,t}/\kappa_t\right)^2 & \left(\kappa_{\mu,t}/\kappa_t\right) \\ \left(\kappa_{\mu,t}/\kappa_t\right) & 1 \end{pmatrix}$$

where $\Sigma^m_{\mu,t|t} = Var\left[\mu \middle| m_{i,t}\right]$. This condition pins down $\kappa_{\mu,t}$ and $\sigma^2_{u,t}$ as a function of κ_t as follows.

$$\kappa_{\mu,t} = \frac{\Sigma^m_{\mu,t|t}}{\Sigma^m_{z,t|t}}\,\kappa_t$$
$$\sigma^2_{u,t} = \kappa_t\left(1-\kappa_t\right)\Sigma^m_{z,t|t} - \kappa_t^2\,\sigma^2_\nu$$

We can see that any $\kappa_t \in \left[0, \frac{\Sigma^m_{z,t|t}}{\Sigma^m_{z,t|t}+\sigma^2_\nu}\right]$ ensures a non-negative $\sigma^2_{u,t}$. Using $e = \begin{pmatrix} 0 & 1 \end{pmatrix}$ to pick out z_t from x_t, we have the following expression for $\tilde{n}_{i,t}$.

$$\tilde{n}_{i,t} = \tilde{K}_t \cdot E\left[x_t \middle| N_t\right] + \tilde{u}_{i,t}, \quad \tilde{u}_{i,t} \sim \mathcal{N}\left(O,\, \sigma_{u,t}\right)$$

where $\tilde{K}_t$ and $\sigma_{u,t}$ are defined as

$$\tilde{K}_t = \kappa_t\,\frac{\Sigma^m_{t|t}\,e\,e'}{e'\,\Sigma^m_{t|t}\,e}$$
$$\sigma_{u,t} = \sigma^2_{u,t}\,\left(e'\,\Sigma^m_{t|t}\,e\right)^{-2}\Sigma^m_{t|t}\,e\,e'\,\Sigma^m_{t|t}$$

for which we use the relationship $\begin{pmatrix} \frac{\kappa_{\mu,t}}{\kappa_t} \\ 1 \end{pmatrix} = \frac{\Sigma^m_{t|t}\,e}{e'\,\Sigma^m_{t|t}\,e}$.

Determination of κ_t We can observe that the optimal κ_t and $\sigma^2_{u,t}$ are equal to the ones determined in Section B. This is because the optimal $n_{i,t}$ under the state vector $x_t = (\mu, z_t)$ is spanned from the optimal $n_{i,t}$ when $x_t = z_t$. The information constraint (4) has the same restriction.

Posterior beliefs It is straightforward to see that the posterior belief evolves as follows, given $\tilde{n}_{i,t}$.

$$x_{i,t|t} = \left(I - \tilde{K}_t\right)x^m_{i,t|t} + \tilde{K}_t\,x_t + \tilde{\nu}_t + \tilde{u}_{i,t}$$
$$\Sigma_{t|t} = \left(I - \tilde{K}_t\right)\Sigma^m_{t|t}$$

where $\tilde{\nu}_t \sim \mathcal{N}\left(O,\, \sigma_\nu\right)$ and $\sigma_\nu = \kappa_t^2\,\left(e'\,\Sigma^m_{t|t}\,e\right)^{-2}\Sigma^m_{t|t}\,e\,e'\,\Sigma^m_{t|t}$.

C.2 Optimal representation of noisy memory

We can express the optimal $m_{i,t}$ as

$$\tilde{m}_{i,t} = \tilde{\Lambda}_t \cdot x_{i,t|t-1} + \tilde{\omega}_{i,t}$$

The feasibility of $\tilde{\Lambda}_t$ is described earlier. The idiosyncratic noise $\tilde{\omega}_{i,t}$ follows a Gaussian distribution $\mathcal{N}(O, \sigma_{\omega,t})$, whose variance is determined by the choice of $\tilde{\Lambda}_t$ as follows.

$$\sigma_{\omega,t} = \left(1 - \tilde{\Lambda}_t\right)\left(Var\left[x_t\right] - \Sigma_{t|t-1}\right)\tilde{\Lambda}_t'$$

Therefore, it remains to specify $\tilde{\Lambda}_t$. The information constraint (6) constrains the choice of $\tilde{\Lambda}_t$. We can derive that

$$\begin{aligned} I\left(m_{i,t}; m_{i,t-1}, n_{i,t-1}\right) &= I\left(\tilde{m}_{i,t}; x_{i,t|t-1}\right) = h\left(\tilde{m}_{i,t}\right) - h\left(\tilde{m}_{i,t} \middle| x_{i,t|t-1}\right) \\ &= \frac{1}{2}\ln\det\left(Var\left[\tilde{m}_{i,t}\right]\right) - \frac{1}{2}\ln\det\left(Var\left[\tilde{m}_{i,t} \middle| x_{i,t|t-1}\right]\right) \\ &= \frac{1}{2}\ln\det\left(Var\left[x_{i,t|t-1}\right]\tilde{\Lambda}_t'\right) - \frac{1}{2}\ln\det\left(\left(I - \tilde{\Lambda}_t\right)Var\left[x_{i,t|t-1}\right]\tilde{\Lambda}_t'\right) \\ &= -\frac{1}{2}\log\det\left(1 - \tilde{\Lambda}_t\right) \leq -\frac{1}{2}\log\,\phi_m \end{aligned}$$

Therefore,

$$\det\left(I - \tilde{\Lambda}_t\right) \geq \phi_m$$

C.2.1 The Choice Variable

Any $\tilde{\Lambda}_t$ is feasible as long as (1) the resulting $\Sigma^m_{t|t}$ is a symmetric and positive semidefinite matrix and (2) the diagonal elements of $\Sigma^m_{t|t}$ are bigger than those of $\Sigma_{t|t-1}$ and smaller than those of σ_x. That is, under any feasible $\tilde{\Lambda}_t$, both $\Sigma^m_{t|t} - \Sigma_{t|t-1}$ and $\sigma_x - \Sigma^m_{t|t}$ are proper variance-covariance matrices (symmetric and positive semidefinite).

It is useful to define $\bar{\Lambda}_t$, which is simply a rotation of $\tilde{\Lambda}_t$.

$$\bar{\Lambda}_t = Var\left[x_{i,t|t-1}\right]^{-\frac{1}{2}}\tilde{\Lambda}_t\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}$$

We could confirm that the same accuracy constraint (6) applies.

$$\begin{aligned} \det\left(I - \bar{\Lambda}_t\right) &= \det\left(I - Var\left[x_{i,t|t-1}\right]^{-\frac{1}{2}}\tilde{\Lambda}_t Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\right) \\ &= \det\left(Var\left[x_{i,t|t-1}\right]^{-\frac{1}{2}}\left(I - \tilde{\Lambda}_t\right)Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\right) = \det\left(I - \tilde{\Lambda}_t\right) \end{aligned}$$

Therefore, I use $W_t = I - \bar{\Lambda}_t$ as a choice variable. Any W_t is feasible as long as W_t and $I - W_t$ are positive semidefinite.

C.2.2 The Constraints

The prior uncertainty is formed according to

$$\begin{aligned} \Sigma^m_{t|t} &= \Sigma_{t|t-1} + \left(I - \tilde{\Lambda}_t\right)Var\left[x_{i,t|t-1}\right] \\ &= \Sigma_{t|t-1} + Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\left(I - \bar{\Lambda}_t\right)Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}} \end{aligned}$$

And the posterior uncertainty can be described as

$$\Sigma_{t|t} = \Sigma_{t|t}^{m} - \left(\kappa_t \, \Sigma_{t|t}^{m} \, e\right) \left(\kappa_t \, e' \, \Sigma_{t|t}^{m} \, e\right)^{-1} \left(\kappa \, e' \, \Sigma_{t|t}^{m}\right)$$
$$= \Sigma_{t|t}^{m} - \Sigma_{t|t}^{m} \, e \, \left(\Omega_{t|t}^{m}\right)^{-1} e' \, \Sigma_{t|t}^{m}$$

where

$$\Omega_{t|t}^{m} = e' \, \Sigma_{t|t}^{m} \, e + \frac{\phi_n}{1-\phi_n} \left(Var\left[z_t\right] + \sigma_\nu\right) + \sigma_\nu^2$$

C.2.3 The Optimization Problem

The optimization problem can then be written as

$$\min_{W_t} tr\left(\sigma_{t|t} \, Q\right)$$

subject to the law of motions of the subjective uncertainty

$$\Sigma_{t|t}^{m} - \Sigma_{t|t-1} = \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \; W_t \; \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}$$
$$\Omega_{t|t}^{m} = e' \, \Sigma_{t|t}^{m} \, e + \frac{\phi_n}{1-\phi_n} \left(Var\left[z_t\right] + \sigma_\nu\right) + \sigma_\nu^2$$
$$\sigma_{t|t} = \Sigma_{t|t}^{m} - \Sigma_{t|t}^{m} \, e \, \left(\Omega_{t|t}^{m}\right)^{-1} e' \, \Sigma_{t|t}^{m}$$

along with the requirement that both W_t and $I - W_t$ are positive semidefinite and symmetric.

Note that when deciding which information to recall at time t (or equivalently, when deciding which information to store at time $t-1$), such a decision takes into account the noisy news that is available at time t. That is, the availability (and the quality) of extra information not from one's memory will affect which information is worthy of remembering. While this is a natural trade-off given the restriction that memory cannot perfectly store all the past information, it is also one that has not been investigated in the literature yet.

C.2.4 Setting up the Lagrange Multipliers

Since W_t is symmetric, it can be eigen-decomposed as $W_t = U \; (I - D) \; U'$ where D is a diagonal matrix and $U \, U' = I$. The constraints that W_t and $I - W_t$ are positive semidefinite are equivalent to the constraints that $I - D$ and D are positive semidefinite. The diagonal elements of $I - D$ and D should be non-negative. The Lagrange multipliers for each inequality constraint can be stored in a diagonal matrix, $\bar{\Upsilon}_1$ and $\bar{\Upsilon}_1$. Finally, I can define $\Upsilon_1 = U \, \bar{\Upsilon}_1 \, U'$ and $\Upsilon_2 = U \, \bar{\Upsilon}_2 \, U'$. Note that $\Upsilon_1 \, W_t = U \, \bar{\Upsilon}_1 \, (I - D) \; U'$ and $\Upsilon_2 \; (I - W_t) = U \, \bar{\Upsilon}_2 \, (D) \; U'$. We can see that the inequality constraint can be expressed as $tr\left(\Upsilon_1 \, W_t\right) \geq 0$ and $tr\left(\Upsilon_2 \; (I - W_t)\right) \geq 0$. This is because $tr\left(\Upsilon_1 \, W_t\right) = tr\left(\bar{\Upsilon}_1 \, (I - D)\right)$ and $tr\left(\Upsilon_2 \; (I - W_t)\right) = tr\left(\bar{\Upsilon}_2 \, (D)\right)$.

We also have equality constraints on the law of motions of subjective uncertainty. For each constraint, I construct a symmetric matrix Γ_i whose kth row contains the Lagrangian multiplier for each kth column of the equality conditions.

C.2.5 The Lagrangian Problem and the First Order Conditions

The Lagrangian problem is as follows.

$$
\begin{aligned}
& max - tr\left(\sigma_{t|t}\, Q\right) \\
& \quad - tr\left(\Gamma_1\left(\left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} W_t \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} + \Sigma_{t|t-1} - \Sigma^m_{t|t}\right)\right) \\
& \quad - tr\left(\Gamma_2\left(e'\,\Sigma^m_{t|t}\, e + \frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma_\nu\right) + \sigma^2_\nu - \Omega^m_{t|t}\right)\right) \\
& \quad - tr\left(\Gamma_3\left(\Sigma^m_{t|t} - \Sigma^m_{t|t}\, e\left(\Omega^m_{t|t}\right)^{-1} e'\,\Sigma^m_{t|t} - \sigma_{t|t}\right)\right) \\
& \quad + tr\left(\Upsilon_1 W_t\right) + tr\left(\Upsilon_2\left(I - W_t\right)\right) + \mu\,\left(\det\left(W_t\right) - \phi_m\right)
\end{aligned}
$$

where the "Langrangian multipliers" Γ_i and Υ_i for all i are symmetric matrices.

The first order conditions subject to W_t, $\Sigma^m_{t|t}$, $\Omega^m_{t|t}$ and $\sigma_{t|t}$ are (in that order)

$$
\begin{aligned}
-\left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \Gamma_1 \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} + \Upsilon_1 - \Upsilon_2 + \mu\,\det\left(W_t\right) W_t^{-1} &= O && \text{(34a)} \\
\Gamma_1 - e\,\Gamma_2\, e' - \Gamma_3 + e\left(\Omega^m_{t|t}\right)^{-1} e'\,\Sigma^m_{t|t}\,\Gamma_3 + \Gamma_3\,\Sigma^m_{t|t}\, e\left(\Omega^m_{t|t}\right)^{-1} e' &= O && \text{(34b)} \\
\Gamma_2 - \left(\Omega^m_{t|t}\right)^{-1} e'\,\Sigma^m_{t|t}\,\Gamma_3\,\Sigma^m_{t|t}\, e\left(\Omega^m_{t|t}\right)^{-1} &= O && \text{(34c)} \\
-Q + \Gamma_3 &= O && \text{(34d)}
\end{aligned}
$$

and the slackness conditions are

$$
\begin{aligned}
\Upsilon_1 W_t = O,\ \Upsilon_1 \succeq O,\ W_t \succeq O && \text{(35a)} \\
\Upsilon_2\left(I - W_t\right) = O,\ \Upsilon_2 \succeq O,\ \left(I - W_t\right) \succeq O && \text{(35b)}
\end{aligned}
$$

and

$$
\mu\,\left(\det\left(W_t\right) - \phi_m\right) = 0,\ \mu \geq 0, \det\left(W_t\right) = \phi_m \qquad \text{(36)}
$$

We can first rearrange (34b)-(34d). Note that $\Gamma_3 = Q$ (as implied by (34d)) and using the notation $\tilde{K}_t \equiv \Sigma^m_{t|t}\, e\left(\Omega^m_{t|t}\right)^{-1} e'$, we can express (34b) as

$$
\Gamma_1 - e\,\Gamma_2\, e' - Q + \tilde{K}_t'\, Q + Q\,\tilde{K}_t = O
$$

and (34c) as

$$
e\,\Gamma_2\, e' - \tilde{K}_t'\, Q\,\tilde{K}_t = O
$$

which together result in

$$\Gamma_1 = \left(I - \tilde{K}_t\right)' Q \left(I - \tilde{K}_t\right)$$

Next, I'd like to solve for W_t that characterizes the optimal memory system. First, multiplying (34a) by $W_t \ (I - W_t)$ on the left yields

$$- \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \Gamma_1 \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} W_t \ (I - W_t) + \mu \, \phi_m \ (I - W_t) = O \tag{37}$$

after applying the slackness conditions (from which $(\Upsilon_1 - \Upsilon_2) \, W_t \ (I - W_t) = O$). We can observe that $\left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \Gamma_1 \left(\sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}$ should be eigen-decomposed in the form of $U \, G \, U'$, that is, it should share the basis with Υ_1, Υ_2 and W_t. Then, the above expression can be written as

$$U \ \left(\mu \, \phi_m \, I - G \ (I - D)\right) D \, U' = O \tag{38}$$

Note that D should satisfy $D \succeq O$, $I - D \succeq O$, and $\det (I - D) = \phi_m$.

C.2.6 The Solution to the Lagrangian Problem

The solution of D can be found as follows. Let's first rearrange U and G so that the diagonal elements in G are in descending order. For $k = 1, \cdots, n$ (where n is the dimension of x_t), I define $\theta_k = \left(\phi_m \prod_{i=1}^{k} g_i\right)^{\frac{1}{k}}$ then we can find k such that $g_k \geq \theta_k > g_{k+1}$ for $k < n$ (or $k = n$ if $g_k \geq \theta_k$). Then, the ith element of D, d_i, is going to be

$$d_i = \begin{cases} 1 - \frac{\theta_k}{g_i} & \text{for } i \leq k \\ 0 & \text{for } i > k \end{cases}$$

We can see that all $d_i \in [0, 1]$ and $\det (I - D) = \Pi_{i=1}^{k} \frac{\theta_k}{g_i} = \phi_m$.

We can express the solution for D more succinctly. Following Afrouzi and Yang (2021), I adopt the following two matrix operators. For a diagonal matrix D, $\max (D, \theta)$ replaces the diagonal elements of D that are smaller than θ with θ. For a symmetric matrix X whose eigendecomposition is expressed as $X = U \, D \, U'$, the operator $Max \, (X, \theta)$ is defined as $Max \, (X, \theta) = U \max (D, \theta) \, U'$. Using these operators, I can express the optimal $I - D$ as

$$I - D = \theta_k \left\{Max \, (G, \theta_k)\right\}^{-1}$$

Since $W_t = U \, (I - D) \, U'$, the optimal solution for W_t is expressed as

$$W_t = \theta_k \ \left\{Max \left(U \, G \, U', \theta_k\right)\right\}^{-1}$$

From this, the optimal $\Sigma_{t|t}^m$ is derived as

$$\Sigma_{t|t}^m = \Sigma_{t|t-1} + Var \left[x_{i,t|t-1}\right]^{\frac{1}{2}} \theta_k \left\{Max \left(Var \left[x_{i,t|t-1}\right]^{\frac{1}{2}} \Gamma_1 \, Var \left[x_{i,t|t-1}\right]^{\frac{1}{2}}, \theta_k\right)\right\}^{-1} Var \left[x_{i,t|t-1}\right]^{\frac{1}{2}}$$

where $Var\left[x_{i,t|t-1}\right]=\sigma_X-\Sigma_{t|t-1}$ captures the maximum possible increase in the uncertainty due to forgetting the previous information $s_{i,t-1}$. In summary, the optimal memory system solves the fixed point problem for Γ_1 and $\Sigma^m_{t|t}$ that satisfy the following equations, given the level of $\Sigma_{t|t-1}$ (and therefore $Var\left[x_{i,t|t-1}\right]$).

$$\Sigma^m_{t|t}=\Sigma_{t|t-1}+Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\theta_k\left\{Max\left(Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\Gamma_1\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}},\theta_k\right)\right\}^{-1}Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}$$
$$\Gamma_1=\left(I-\tilde{K}_t\right)'\,Q\,\left(I-\tilde{K}_t\right)$$

Furthermore, as summarized by $\tilde{\Lambda}_t$, the optimal memory signal is described as follows.

$$\tilde{\Lambda}_t=Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\left(\sum_{i=1}^{k}\left(1-\frac{\theta_k}{g_i}\right)u_i\,u_i'\right)Var\left[x_{i,t|t-1}\right]^{-\frac{1}{2}}$$

where g_i is the eigenvalues of $Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\Gamma_1\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}$ that are rearranged in a descending order and u_i is the corresponding eigenvector. As defined above, k is such that $g_k\geq\theta_k\geq g_{k+1}$.

D Derivations of β_I and β_C (when the long-run mean is known)

DM i's forecast of z_t evolves according to the following linear law of motion.

$$z_{i,t|t} = (1-\lambda)(1-\kappa)\,\mu + \lambda\,(1-\kappa)\,z_{i,t|t-1} + \kappa\,z_t + \kappa\,\tilde{\nu}_t + \tilde{u}_{i,t} + (1-\kappa)\,\tilde{\omega}_{i,t}$$

The consensus forecast of z_t evolves according to the following linear law of motion.

$$z_{t|t} = (1-\lambda)(1-\kappa)\,\mu + \lambda\,(1-\kappa)\,z_{t|t-1} + \kappa\,z_t + \kappa\,\tilde{\nu}_t \tag{39}$$

I define b as the weight on unconditional prior belief.

$$b \equiv (1-\lambda)(1-\kappa) \tag{40}$$

D.1 Derivations of β_I and β_C

Derivation of β_I

From the regression specification

$$z_t - z_{i,t|t} = \alpha_I + \beta_I\left(z_{i,t|t} - z_{i,t|t-1}\right) + error_{i,t},$$

the coefficient β_I asymptotically converges to

$$\beta_I = \frac{Cov\left[z_t - z_{i,t|t},\, z_{i,t|t} - z_{i,t|t-1}\right]}{Var\left[z_{i,t|t} - z_{i,t|t-1}\right]}$$

We can see that

$$Cov\left[z_t - z_{i,t|t},\, z_{i,t|t} - z_{i,t|t-1}\right] = -Cov\left[z_t - z_{i,t|t},\, z_{i,t|t-1}\right] = -b\,Var\left[z_{i,t|t}\right]$$

The first equality holds because $Cov\left[z_t - z_{i,t|t},\, z_{i,t|t}\right] = 0$. The second equality holds because $E\left[z_{i,t|t}\middle|\, m_{i,t-1}, n_{i,t-1}\right] = b\,\mu + (1-b)\,z_{i,t|t-1}$. We can also see that

$$Var\left[z_{i,t|t} - z_{i,t|t-1}\right] = \left(\rho^{-2} - 2\,(1-b) + 1\right)Var\left[z_{i,t|t-1}\right]$$

where I use $Var\left[z_{i,t|t-1}\right] = \rho^2\,Var\left[z_{i,t|t}\right]$. Combining the two derivations, we get

$$\beta_I = -\frac{b}{2\,b + \rho^{-2} - 1} \tag{41}$$

Derivation of β_C

Rearranging terms, we can express the consensus forecast's error as follows.

$$z_t - z_{t|t} = \frac{1-\kappa}{\kappa}\left(z_{t|t} - z_{t|t-1} + (1-\lambda)\left(z_{t|t-1} - \mu\right)\right) - \tilde{\nu}_t$$

From the regression specification

$$z_t - z_{t|t} = \alpha_C + \beta_C \left(z_{t|t} - z_{t|t-1}\right) + error_t,$$

the coefficient β_C asymptotically converges to

$$\beta_C = \frac{Cov\left[z_t - z_{t|t}, z_{t|t} - z_{t|t-1}\right]}{Var\left[z_{t|t} - z_{t|t-1}\right]}$$

Therefore, we can see that

$$\beta_C = \frac{1-\kappa}{\kappa}\left(1 + (1-\lambda)\frac{Cov\left[z_{t|t-1}, z_{t|t} - z_{t|t-1}\right]}{Var\left[z_{t|t} - z_{t|t-1}\right]}\right) - \frac{\kappa\,\sigma_\nu^2}{Var\left[z_{t|t} - z_{t|t-1}\right]}$$

It remains to derive expressions for $Cov\left[z_{t|t-1}, z_{t|t} - z_{t|t-1}\right]$ and $Var\left[z_{t|t} - z_{t|t-1}\right]$.

Note that

$$\begin{aligned}(1-\lambda(1-\kappa)\rho L)\, z_{t|t} &= \kappa\left(z_t + \tilde{\nu}_t\right)\\ \Leftrightarrow \quad z_{t|t} &= \frac{\kappa}{1-\lambda(1-\kappa)\rho L}\left(z_t + \tilde{\nu}_t\right)\end{aligned}$$

Therefore, it is straightforward to see that

$$Cov\left[z_t, z_{t|t}\right] = \frac{\kappa}{1-\lambda(1-\kappa)\rho^2} Var\left[z_t\right]$$

We can also show that

$$\begin{aligned}Var\left[z_{t|t}\right] &= Var\left[\frac{\kappa}{1-\lambda(1-\kappa)\rho L}\frac{1}{1-\rho L}\epsilon_t + \frac{\kappa}{1-\lambda(1-\kappa)\rho L}\tilde{\nu}_t\right]\\ &= \left[\frac{1+\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\frac{\kappa^2}{1-(\lambda(1-\kappa)\rho)^2}\frac{\sigma_\epsilon^2}{1-\rho^2}\right] + \left[\frac{\kappa^2}{1-(\lambda(1-\kappa)\rho)^2}\sigma_\nu^2\right]\\ &= \frac{\kappa^2}{1-(\lambda(1-\kappa)\rho)^2}\left\{\frac{1+\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2} Var\left[z_t\right] + \sigma_\nu^2\right\}\end{aligned}$$

And finally,

$$Cov\left[z_{t|t}, z_{t|t-1}\right] = \lambda(1-\kappa)\rho^2\, Var\left[z_{t|t}\right] + \kappa\,\rho^2\, Cov\left[z_t, z_{t|t}\right]$$

Let's consider the case $\sigma_\nu^2 \to 0$. Then,

$$\begin{aligned}Cov\left[z_t, z_{t|t}\right] &= \frac{1}{k}\frac{1-(\lambda(1-\kappa)\rho)^2}{1+\lambda(1-\kappa)\rho^2} Var\left[z_{t|t}\right]\\ Cov\left[z_{t|t}, z_{t|t-1}\right] &= \left[\kappa\,\rho^2 + \kappa\,\rho^2\,\frac{1}{k}\frac{1-(\lambda(1-\kappa)\rho)^2}{1+\lambda(1-\kappa)\rho^2}\right] Var\left[z_{t|t}\right]\\ &= \frac{\rho^2 + \lambda(1-\kappa)\rho^2}{1+\lambda(1-\kappa)\rho^2} Var\left[z_{t|t}\right] \equiv \bar{c}\, Var\left[z_{t|t}\right]\end{aligned}$$

Then,

$$\frac{Cov\left[z_{t|t-1}, z_{t|t}-z_{t|t-1}\right]}{Var\left[z_{t|t}-z_{t|t-1}\right]} = \frac{\left(\bar{c}-\rho^2\right)\, Var\left[z_{t|t}\right]}{\left(1+\rho^2-2\,\bar{c}\right) Var\left[z_{t|t}\right]} = \frac{\bar{c}-\rho^2}{1+\rho^2-2\,\bar{c}} = \frac{\lambda\left(1-\kappa\right)\rho^2}{1-\lambda\left(1-\kappa\right)\rho^2}$$

Finally, we can derive that β_C is expressed as follows.

$$\beta_C = \frac{1-\kappa}{\kappa}\left(1+\left(1-\lambda\right)\frac{\lambda\left(1-\kappa\right)\rho^2}{1-\lambda\left(1-\kappa\right)\rho^2}\right) \tag{42}$$

D.2 Steady-state Uncertainty

I denote the steady state uncertainty of z_t as $\Sigma_{-1} \equiv Var\left[z_t|\, m_{i,t-1}, n_{i,t-1}\right]$, $\Sigma^m \equiv Var\left[z_t|\, m_{i,t}\right]$, and $\Sigma \equiv Var\left[z_t|\, m_{i,t}, n_{i,t}\right]$, which satisfy the following stationary relationship.

$$\Sigma_{-1} = \rho^2\,\Sigma + \sigma_\epsilon^2 \tag{43a}$$

$$\Sigma^m = \left(1-\lambda\right)\sigma_z^2 + \lambda\,\Sigma_{-1} \tag{43b}$$

$$\left(\Sigma\right)^{-1} = \left(\Sigma^m\right)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} \tag{43c}$$

where σ_z^2 is the unconditional variance of z, which equals $= \frac{\sigma_\epsilon^2}{1-\rho^2}$, and $\tilde{\sigma}_u^2 = \frac{\phi_n}{1-\phi_n}\,\sigma_z^2$ captures the noisy news.

The steady-state κ and b are

$$\kappa = \frac{\Sigma^m}{\Sigma^m+\tilde{\sigma}_u^2} \tag{44}$$

$$b = \left(1-\lambda\right)\,\frac{\tilde{\sigma}_u^2}{\Sigma^m+\tilde{\sigma}_u^2} \tag{45}$$

And we have shown earlier that $\lambda = 1-\phi_m$.

D.3 Comparative Statics

Comparative Statics for the Uncertainty

Equations (43) implicitly impose the following relation.

$$\begin{aligned} F\left(\Sigma;\tilde{\sigma}_u^2,\lambda\right) &= \left(\Sigma^m\right)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - \left(\Sigma\right)^{-1} \\ &= \left(\left(1-\lambda\right)\sigma_z^2 + \lambda\,\left(\rho^2\,\Sigma+\sigma_\epsilon^2\right)\right)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - \left(\Sigma\right)^{-1} = 0 \end{aligned} \tag{46}$$

Then, the derivatives of $F\left(\Sigma;\tilde{\sigma}_u^2,\lambda\right)=0$ with respect to $\tilde{\sigma}_u^2$ and λ are

$$\frac{\partial F}{\partial \tilde{\sigma}_u^2} = -\left(\Sigma^m\right)^{-2}\lambda\,\rho^2\,\frac{\partial \Sigma}{\partial \tilde{\sigma}_u^2} - \left(\tilde{\sigma}_u^2\right)^{-2} + \left(\Sigma\right)^{-2}\frac{\partial \Sigma}{\partial \tilde{\sigma}_u^2} = 0$$

$$\frac{\partial F}{\partial \lambda} = -\left(\Sigma^m\right)^{-2}\left(-\sigma_z^2 + \rho^2\Sigma + \sigma_\epsilon^2 + \lambda\,\rho^2\,\frac{\partial \Sigma}{\partial \lambda}\right) + \left(\Sigma\right)^{-2}\frac{\partial \Sigma}{\partial \lambda} = 0$$

Rearranging yields the derivatives of σ with respect to $\tilde{\sigma}_u^2$ and λ.

$$\frac{\partial \Sigma}{\partial \tilde{\sigma}_u^2} = \left(\left(\frac{\Sigma^m}{\Sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \left(\frac{\Sigma^m}{\tilde{\sigma}_u^2} \right)^2 > 0$$

$$\begin{aligned}
\frac{\partial \Sigma}{\partial \lambda} &= - \left(\left(\frac{\Sigma^m}{\Sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \left(\sigma_z^2 - \Sigma_{-1} \right) \\
&= - \left(\left(\frac{\Sigma^m}{\sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \frac{\Sigma^m}{1-\lambda} \left(1 - \frac{\Sigma_{-1}}{\Sigma^m} \right) < 0
\end{aligned}$$

Additionally, the derivative of Σ^m with respect to $\tilde{\sigma}_u^2$ is

$$\frac{\partial \Sigma^m}{\partial \tilde{\sigma}_u^2} = \lambda\, \rho^2\, \frac{\partial \Sigma}{\partial \tilde{\sigma}_u^2} = \lambda\, \rho^2 \left(\left(\frac{\Sigma^m}{\Sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \left(\frac{\Sigma^m}{\tilde{\sigma}_u^2} \right)^2 > 0$$

and with respect to λ:

$$\begin{aligned}
\frac{\partial \Sigma^m}{\partial \lambda} &= -\rho^2 \left(\sigma_z^2 - \Sigma \right) + \lambda\, \rho^2\, \frac{\partial \Sigma}{\partial \lambda} \\
&= - \frac{\Sigma^m}{1-\lambda} \left(1 - \frac{\Sigma_{-1}}{\Sigma^m} \right) \left\{ 1 + \frac{\lambda\, \rho^2}{\left(\frac{\Sigma^m}{\Sigma}\right)^2 - \lambda \rho^2} \right\} \\
&= - \frac{\Sigma^m}{1-\lambda} \frac{1 - \frac{\Sigma_{-1}}{\Sigma^m}}{1 - \lambda \rho^2\, \left(\frac{\Sigma}{\Sigma^m}\right)^2} < 0
\end{aligned}$$

Note that $1 > \frac{\Sigma_{-1}}{\Sigma^m} > \frac{\Sigma}{\Sigma^m} > \lambda \rho^2\, \left(\frac{\Sigma}{\Sigma^m}\right)^2 > 0$, making the last term be between 0 and 1.

Comparative Statics for κ and b

Now we turn to the comparative statistics of κ and b. First, the derivative of b with respect to $\tilde{\sigma}_u^2$ is computed as:

$$\begin{aligned}
\frac{\partial b}{\partial \tilde{\sigma}_u^2} &= (1-\lambda)\, \frac{1}{(\Sigma^m + \tilde{\sigma}_u^2)^2} \left\{ \left(\Sigma^m + \tilde{\sigma}_u^2 \right) - \tilde{\sigma}_u^2 \left(\frac{\partial \Sigma^m}{\partial \tilde{\sigma}_u^2} + 1 \right) \right\} \\
&= (1-\lambda)\, \frac{\Sigma^m}{(\Sigma^m + \tilde{\sigma}_u^2)^2} \left\{ 1 - \lambda\, \rho^2\, \frac{\frac{\Sigma^m}{\Sigma} - 1}{\left(\frac{\Sigma^m}{\Sigma}\right)^2 - \lambda \rho^2} \right\} > 0
\end{aligned}$$

We can easily see that $\frac{\frac{\Sigma^m}{\Sigma} - 1}{\left(\frac{\Sigma^m}{\Sigma}\right)^2 - \lambda \rho^2} \in (0,1)$, which makes the term inside the bracket be positive. Next, the derivative of b with respect to λ is derived as:

$$\begin{aligned}
\frac{\partial b}{\partial \lambda} &= - \frac{\tilde{\sigma}_u^2}{\Sigma_m + \tilde{\sigma}_u^2} - (1-\lambda)\, \frac{\tilde{\sigma}_u^2}{(\Sigma_m + \tilde{\sigma}_u^2)^2} \frac{\partial \Sigma^m}{\partial \lambda} = - \frac{\tilde{\sigma}_u^2}{\Sigma_m + \tilde{\sigma}_u^2} \left(1 + \frac{1-\lambda}{\Sigma_m + \tilde{\sigma}_u^2} \frac{\partial \Sigma^m}{\partial \lambda} \right) \\
&= - \frac{\tilde{\sigma}_u^2}{\Sigma_m + \tilde{\sigma}_u^2} \left(1 - \frac{\Sigma^m}{\Sigma_m + \tilde{\sigma}_u^2} \frac{1 - \frac{\Sigma_{-1}}{\Sigma^m}}{1 - \lambda \rho^2\, \left(\frac{\Sigma}{\Sigma^m}\right)^2} \right) < 0
\end{aligned}$$

In addition, the derivative of κ with respect to $\tilde{\sigma}_u^2$ is:

$$\begin{aligned}\frac{\partial \kappa}{\partial \tilde{\sigma}_u^2} &= -\frac{\Sigma^m}{(\Sigma^m + \tilde{\sigma}_u^2)^2}\left\{1 - \frac{\partial \Sigma^m}{\partial \tilde{\sigma}_u^2}\frac{\tilde{\sigma}_u^2}{\Sigma^m}\right\} \\ &= -\frac{\Sigma^m}{(\Sigma^m + \tilde{\sigma}_u^2)^2}\left\{1 - \lambda\rho^2\frac{\frac{\Sigma^m}{\Sigma} - 1}{\left(\frac{\Sigma^m}{\Sigma}\right)^2 - \lambda\rho^2}\right\} < 0\end{aligned}$$

Finally, the derivative of κ with respect to λ:

$$\frac{\partial \kappa}{\partial \lambda} = \frac{\tilde{\sigma}_u^2}{(\Sigma^m + \tilde{\sigma}_u^2)^2}\frac{\partial \Sigma^m}{\partial \lambda} < 0$$

Comparative Statics for β_I

Now we combine the above comparative statistics to analyze how β_I and β_C change with ϕ_n and ϕ_m. Note first from (41) that ϕ_n and ϕ_m affect β_I through the bias term b. The derivative of β_I with respect to b is:

$$\frac{\partial \beta_I}{\partial b} = -\left(2\,b + \rho^{-2} - 1\right)^{-2}\left(\rho^2 - 1\right) < 0$$

Therefore, we get that

$$\frac{\partial \beta_I}{\partial \phi_m} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \phi_m} = -\frac{\partial \beta_I}{\partial b}\frac{\partial \beta_I}{\partial \lambda} < 0 \tag{47a}$$

$$\frac{\partial \beta_I}{\partial \phi_n} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} < 0 \tag{47b}$$

Comparative Statics for β_C

Next, we analyze the comparative statics for β_C. First, the derivative of β_C with respect to ϕ_n is more straightforward. From (42), we can see that β_C decreases in κ, and from above we also know that κ decreases in $\tilde{\sigma}_u^2$. Therefore, we have

$$\frac{\partial \beta_C}{\partial \phi_n} = \frac{\partial \beta_C}{\partial \kappa}\frac{\partial \kappa}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} > 0 \tag{48}$$

The derivative of β_C with respect to ϕ_m is more involved. We can compute that

$$
\begin{aligned}
\frac{\partial \beta_C}{\partial \phi_m} &= -\frac{1}{\kappa^2}\frac{\partial \kappa}{\partial \phi_m}\left(1+(1-\lambda)\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)+\frac{1-\kappa}{\kappa}\frac{\partial}{\partial \phi_m}\left(1+(1-\lambda)\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\\
&= -\frac{1}{\kappa^2}\frac{\partial \kappa}{\partial \phi_m}\left(1+(1-\lambda)\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\\
&\quad+\frac{1-\kappa}{\kappa}(1-\lambda)\frac{\partial}{\partial \phi_m}\left(\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)+\frac{1-\kappa}{\kappa}\left(\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\frac{\partial(1-\lambda)}{\partial \phi_m}\\
&= \underbrace{-\frac{1}{\kappa^2}\frac{\partial \kappa}{\partial \phi_m}}_{<0}+\underbrace{\frac{1-\kappa}{\kappa}(1-\lambda)\frac{\partial}{\partial \phi_m}\left(\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)}_{<0}-\frac{1}{\kappa^2}\left(\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\left\{(1-\lambda)\underbrace{\frac{\partial \kappa}{\partial \phi_m}}_{>0}-k(1-\kappa)\right\}
\end{aligned}
$$

The last equation holds because $\frac{\partial(1-\lambda)}{\partial \phi_m}=1$. Since all the terms except the last one are negatively contributing to $\frac{\partial \beta_C}{\partial \phi_m}$, we can further see that

$$
\begin{aligned}
\frac{\partial \beta_C}{\partial \phi_m} &< -\frac{1}{\kappa^2}\frac{\partial \kappa}{\partial \phi_m}+\frac{1-\kappa}{\kappa}\left(\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\\
&= -\frac{1-\kappa}{\kappa}\left(\frac{\partial \sigma^m}{\partial \phi_m}-\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\\
&= -\frac{1-\kappa}{\kappa}\left(\frac{(1-\kappa)\rho^2\left(\frac{\sigma_z^2}{\sigma}-1\right)}{1-\lambda(1-\kappa)^2\rho^2}-\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\right)\\
&= -\frac{1-\kappa}{\kappa}\frac{\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)\rho^2}\left\{\frac{1}{\lambda}\left(\frac{\sigma_z^2}{\Sigma}-1\right)\frac{1-\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)^2\rho^2}-1\right\}
\end{aligned}
$$

I would like to show that we can find $\hat{\sigma}_u^2$ such that for any λ, the term in the bracket is positive for all $\tilde{\sigma}_u^2$ such that $\tilde{\sigma}_u^2 \le \hat{\sigma}_u^2$ and negative otherwise.

First, it is straightforward to see that the term in the bracket is positive for $\tilde{\sigma}_u^2=0$ (since $\Sigma \to 0$) and negative for $\tilde{\sigma}_u^2 \to \infty$ (since $\sigma \to \sigma_z^2$) for any values of ρ, σ_ϵ, and λ. Next, we can also see that the term in the bracket is decreasing in $\tilde{\sigma}_u^2$ for any given ρ, σ_ϵ, and λ: $\frac{\sigma_z^2}{\sigma}$ decreases in $\tilde{\sigma}_u^2$ and $\frac{1-\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)^2\rho^2}$ decreases in $1-k$ (and accordingly also decreases in $\tilde{\sigma}_u^2$). Therefore, there exists a $\hat{\sigma}_u^2$ such that the term in the bracket is positive for any ρ, σ_ϵ, and λ as long as $\sigma_u \le \hat{\sigma}_u^2$. In practice, we could find such $\hat{\sigma}_u^2$ by finding $\tilde{\sigma}_u^2$ under which

$$
\frac{1}{\lambda}\left(\frac{\sigma_z^2}{\Sigma}-1\right)\frac{1-\lambda(1-\kappa)\rho^2}{1-\lambda(1-\kappa)^2\rho^2}=1
$$

for any given ρ, σ_ϵ^2 and λ. For a given value of ρ and σ_ϵ^2, we can define the minimum $\hat{\sigma}_u^2$ for all λ as $\hat{\sigma}_u^2 \equiv g(\rho,\sigma_\epsilon)$. Therefore, we can conclude that $\frac{\partial \beta_C}{\partial \phi_m}<0$ as long as $\tilde{\sigma}_u^2 \le g\left(\rho,\sigma_\epsilon^2\right)$. Equivalently, $\frac{\partial \beta_C}{\partial \phi_m}<0$ as long as $\phi_n \le \bar{\phi}_n \equiv \bar{g}\left(\rho,\sigma_\epsilon^2\right)$, where $\bar{g}\left(\rho,\sigma_\epsilon^2\right)$ can be easily defined using the definition $\tilde{\sigma}_u^2=\frac{\phi_n}{1-\phi_n}\sigma_z^2$.

E Estimation

E.1 Data Source Description

Survey Forecasts Data

The Survey of Professional Forecasters (SPF) began in 1968:Q4 and was taken over by the Philadelphia Fed in 1990:Q2. Forecasters submit their projections in the middle month of each quarter. Two major new data releases are available to the survey participants before submitting their survey. One is the release of the Bureau of Economic Analysis' advance report of the national income and product accounts, which contains the first estimate of GDP and its components for the previous quarter. This is released at the end of the first month of each quarter. The other is the release of the Bureau of Labor Statistics' monthly Employment Situation Report, which is released on the first Friday of each month.

Variable information I use the following eleven variables in the "U.S. Business Indicators" Section. To ease the notation burden, I use the acronym when necessary.

1. Nominal Gross Domestic Product (NGDP)
    - Seasonally adjusted, annual rate
    - Before 1992, forecasts for nominal GNP
2. Real Gross Domestic Product (RGDP)
    - Seasonally adjusted, annual rate
    - Chain-weighted real GDP. Before 1992, fixed-weighted real GDP. Before 1981:Q3, RGDP is computed as NGDP/PGDP*100.
3. GDP Chain-Weighted Price Index (PGDP)
    - Seasonally adjusted, annual rate
    - Chain-weighted GDP price index. The base year varies. Before 1992, GNP deflator.
4. Corporate Profits After Taxes (CPROF)
    - Seasonally adjusted, annual rate
    - Before 2006, nominal corporate profits after tax, excluding inventory valuation adjustment (IVA) and capital consumption adjustment (CCAdj)
5. Civilian Unemployment Rate (UNEMP)
    - Seasonally adjusted
    - Quarterly average of the monthly unemployment rates

6. Industrial Production Index (INDPROD)
    - Seasonally adjusted
    - The base year of the index varies
    - Quarterly average of the monthly levels
7. Housing Starts (HOUSING)
    - Seasonally adjusted, annual rate
    - Quarterly average of the monthly levels
8. Consumer Price Index (CPI)
    - Seasonally adjusted
    - Headline CPI inflation rate. The unit of the quarterly forecasts is a quarter-over-quarter annualized growth rate of the quarterly average price index level
    - Survey starts in 1981:Q3
9. 3-month Treasury Bill Rate (TBILL)
    - Quarterly average of the daily levels
10. AAA Corporate Bond Yield (BOND)
    - Quarterly average of the daily levels of Moody's Aaa corporate bond yields
    - Before 1990Q4, new, high-grade corporate bond yield
11. 10-year Treasury Bond Rate (TBOND)
    - Quarterly average of the daily levels of 10-year Treasury bond rate

Data availability The survey forecasts have been available for most of these variables since 1968Q4. Exceptions are CPI, TBILL, BOND, and TBOND, whose survey forecasts became available in 1981.

Forecast horizons Forecasters provide (1) quarterly projections for five quarters (current and up to four-quarter-ahead) and (2) annual projections for the current and the following year. For this paper, I use quarterly projections.

Forecast unit Forecasters could provide forecasts using either level or growth rates for most variables. The exception is the forecasts for CPI and PCE, for which forecasters make quarter-over-quarter forecasts.

I compute forecasters' projections about how the variables will change from the previous quarter. For most variables, I take a log difference. For the financial variables and the unemployment, I take the difference. I annualize this difference to compare across different forecast horizons. For example, for the variables I take the log-difference, forecasts are defined as

$$F_{i,t}\, y_{t+h} = \left(\log\left(F_{i,t}\, Y_{t+h}\right) - \log\left(F_{i,t}\, Y_{t-1}\right)\right) \times \frac{4}{h}$$
$$F_{i,t-1}\, y_{t+h} = \left(\log\left(F_{i,t-1}\, Y_{t+h}\right) - \log\left(F_{i,t-1}\, Y_{t-1}\right)\right) \times \frac{4}{h+1}$$

For variables I take the difference, forecasts are defined as

$$F_{i,t}\, y_{t+h} = \left(F_{i,t}\, Y_{t+h} - F_{i,t}\, Y_{t-1}\right) \times \frac{4}{h}$$
$$F_{i,t-1}\, y_{t+h} = \left(F_{i,t-1}\, Y_{t+h} - F_{i,t-1}\, Y_{t-1}\right) \times \frac{4}{h+1}$$

When computing the forecast revision, I compare these forecasts to those made in the previous quarter. Forecasts from the previous quarters are projections about how the variables will change in the next quarter. Forecast revisions are defined as

$$F_{i,t}\, y_{t+h} - F_{i,t-1}\, y_{t+h}$$

Outlier treatment After constructing the forecasts described above, I drop some observations to restrict the influence of a few outlier variables. First, in each period, I remove forecasts that are five quantiles outsides of the median forecasts. And I only keep individual forecasts that have more than ten observations of the error-revision pairs.

I further restrict samples to measure the forecast behavior in the normal business cycle. During a likely structural change, forecasters might use different forecasting models than the one they would use during the regular cycle. To systematically identify these episodes, I compute the average size of forecast revisions among forecasters each period and remove the top 5 percentile periods. I find such periods of extensive revisions for each forecast horizon. For variables of 200-period observations, I am dropping ten periods. For example, here is the list of periods removed for the forecast of the current quarter realizations for each variable.

1. NGDP: 1974q4, 1975q1, 1980q1, 1981q3, 1981q4, 2001q4, 2008q4, 2009q1, 2020q2, 2020q3

2. RGDP: 1970q4, 1974q4, 1975q1, 1980q1, 1980q2, 1981q4, 2001q4, 2009q1, 2020q2, 2020q3

3. PGDP: 1970q2, 1973q4, 1974q2, 1974q4, 1975q2, 1979q3, 1980q2, 1981q3, 2020q2, 2022q2

4. CPROF: 1974q4, 1981q1, 1981q4, 1982q1, 1982q2, 2002q1, 2005q4, 2020q2, 2020q3 2020q4

5. UNEMP 1974q4, 1975q1, 1980q2, 1981q4, 1982q4, 2001q4, 2009q1, 2009q2, 2020q2, 2020q3, 2020q4
6. INDPROD: 1970q4, 1974q4, 1975q1, 1980q2, 1980q3, 1981q4, 1982q1, 1982q4, 2020q2, 2020q3
7. HOUSING: 1973q4, 1974q4, 1978q2, 1980q2, 1981q1, 1981q3, 1981q4, 2009q1, 2020q2, 2020q3
8. CPI: 1982q1, 1983q1, 1986q2, 1990q4, 2008q4, 2009q1, 2015q1, 2020q2
9. TBILL: 1981q4, 1982q1, 1982q3, 1982q4, 1984q4, 2001q4, 2008q4, 2020q2
10. BOND: 1981q4, 1982q1, 1982q2, 1982q3, 1982q4, 1983q3, 1984q2, 1994q2
11. TBOND: 1992q3, 1994q2, 1996q2, 2002q3, 2008q1, 2020q2, 2022q2

Real-time Macroeconomic Data

I use the real-time data set provided by the Philadelphia Fed. The first release of each variable is used as the "true" realization, which has two uses for my exercise. First, I use this data to compute the forecast errors. Second, I estimate the parameters related to the data-generating process using this data. The last data point I use is 2019:Q4. This is because many variables have an abrupt change during the Covid period, which I assume is not well described as a stationary distribution.

Using real-time data allows us to compute the forecast error correctly. Macroeconomic variables are redefined or reclassified, and the base year changes for the real variables. Therefore, we must compare the forecast data to a correct realized macro variable with a consistent definition. The real-time data includes the latest data available at any given vintage. The data released for the same vintage is constructed based on an internally consistent variable definition and the same base year. At least for data released after 1996 (when the chain weighting replaced the fixed-weighing method), the change of base year doesn't affect the growth rate of variables.

E.2 Regression Estimation

As discussed in the main text, I estimate three regressions. First, following the specification proposed in Bordalo, Gennaioli, Ma, and Shleifer (2020b), I estimate the following regression.

$$y_{t+h} - F_{i,t}\, y_{t+h} = \alpha_{i,I} + \beta_I\, \left(F_{i,t}\, y_{t+h} - F_{i,t-1}\, y_{t+h}\right) + I_h + error_{i,t,h} \tag{49}$$

$F_{i,t}\, y_{t+h}$ is forecaster i's projected h-quarter-ahead change of y_t from the previous quarter, and the revision variable captures how her belief changed from the previous quarter. $\alpha_{i,I}$ is a dummy variable for each forecaster, and I_h is a dummy variable for each forecast horizon that ranges from $h = 0$ to $h = 3$. I pool all forecast horizons when estimating β_I. The top panel in Table 7 reports the results.

The second regression is from Coibion and Gorodnichenko (2015).

$$y_{t+h} - F_t\, y_{t+h} = \alpha_C + \beta_C\, (F_t\, y_{t+h} - F_{t-1}\, y_{t+h}) + I_h + error_{t,h} \tag{50}$$

$F_t\, y_{t+h}$ is the average forecast of $F_{i,t}\, y_{t+h}$, for which I use the sample mean of individual forecasts at any given time t. Again, I pool all forecast horizons when estimating β_C. The middle panel in Table 7 reports the results.

Finally, the last regression follows the specification from Gemmi and Valchev (2021).

$$F_{i,t}\, y_{t+h} - F_{i,t-1}\, y_{t+h} = \alpha_{i,K} + \beta_K\, (F_{t-1}\, y_{t+h} - F_{i,t-1}\, y_{t+h}) + D_t + I_h + error_{i,t} \tag{51}$$

where D_t is the time dummy. This specification, in essence, regresses the de-meaned forecast revision on de-meaned forecast surprises (defined as $y_{t+h} - F_{i,t-1}\, y_{t+h}$). All forecast horizons are pooled. The bottom panel in Table 7 reports the results.

I also report the estimated regression coefficients using only a single forecast horizon. Table 8 shows the coefficients estimated from the current quarter forecasts. And Table 9 shows the coefficients estimated from the three-quarter-ahead forecasts. Finally, I also report the coefficients using the entire sample period in Table 10. For this version, I do not drop the high-mean-squared-error periods identified in the previous section.

CPI long-term forecasts

To estimate the uncertainty about the long-run mean, I estimate how forecasts of μ are revised in response to news about the current quarter. I use the following specification to build on the intuition of Gemmi and Valchev (2021).

$$(F_{i,t}\, \mu - F_{i,t-1}\, \mu) - (F_t\, \mu - F_{t-1}\, \mu) = \alpha_{i,\mu,K} + \beta_{\mu,K}\, (F_{t-1}\, \mu - F_{i,t-1}\, \mu) + error_{i,\mu,t} \tag{52}$$

where $F_{i,t}\, \mu$ is the forecast about the long-run, and $F_{i,t}\, \mu$ is the average of $F_{i,t}\, \mu$ across forecasters at time t.

Among the forecast data, the only variable that allows the estimation of the above regression specification is that of the CPI. SPF asks panelists to submit their views about the annual average rate of headline CPI inflation over the next five and ten years. The five-year forecast data started in 2005Q3, and the ten-year forecast data started in 1991Q4. Table 11 reports the estimation results. I also report the response of the three-quarter-ahead and the current-quarter forecasts in response to the news about the current quarter's CPI as a comparison. Unlike the previous regression specifications in Table 7, I transform the quarterly forecast data to reflect the annual average inflation rate to maintain the definition consistent with the long-term forecast data.

Table 7: Baseline Regression Coefficients

(a) Bordalo et al. Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	-0.271***	-0.249***	-0.318***	-0.440***	-0.0476	-0.180***	-0.252***	-0.174**	-0.00757	-0.349***	-0.124**
	(0.0446)	(0.0515)	(0.0445)	(0.0465)	(0.0470)	(0.0607)	(0.0603)	(0.0853)	(0.0390)	(0.0438)	(0.0478)
N	20919	20875	20657	14646	21279	19364	20126	14722	14993	12551	12645

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(b) Coibion-Gorodnichenko Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	0.632***	0.452***	0.706***	0.685***	0.617***	0.610***	0.505***	0.548***	0.294***	0.0528	0.283***
	(0.106)	(0.129)	(0.129)	(0.181)	(0.0913)	(0.139)	(0.113)	(0.164)	(0.0459)	(0.0719)	(0.0573)
N	797	796	799	791	800	796	795	604	614	613	455

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are robust to arbitrary heteroskedasticity and autocorrelation.

Dummy variables for forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(c) Gemmi-Valchev Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Surprise	0.545***	0.600***	0.600***	0.513***	0.623***	0.576***	0.579***	0.560***	0.727***	0.678***	0.701***
	(0.0256)	(0.0227)	(0.0277)	(0.0289)	(0.0190)	(0.0239)	(0.0241)	(0.0303)	(0.0319)	(0.0234)	(0.0275)
N	21302	21268	20950	14896	21560	19684	20463	15008	15157	12638	12727

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for time, forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

Table 8: Regression Coefficients for Current-quarter Forecasts Only

(a) Bordalo et al. Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	-0.300***	-0.310***	-0.365***	-0.438***	-0.0794*	-0.193***	-0.292***	-0.0795	-0.0742**	-0.327***	-0.0546
	(0.0473)	(0.0536)	(0.0508)	(0.0618)	(0.0448)	(0.0684)	(0.0609)	(0.0762)	(0.0370)	(0.0412)	(0.0394)
N	5346	5357	5289	3722	5305	4918	5155	3757	3794	3183	3175

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(b) Coibion-Gorodnichenko Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	0.524***	0.243	0.553***	0.879***	0.455***	0.483**	0.369***	0.567**	0.182***	0.0351	0.293***
	(0.159)	(0.185)	(0.199)	(0.264)	(0.0892)	(0.190)	(0.139)	(0.229)	(0.0447)	(0.0921)	(0.0587)
N	202	202	203	201	202	202	202	153	155	155	114

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are robust to arbitrary heteroskedasticity and autocorrelation.

Dummy variables for forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(c) Gemmi-Valchev Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Surprise	0.591***	0.669***	0.666***	0.533***	0.756***	0.669***	0.673***	0.653***	0.921***	0.851***	0.861***
	(0.0312)	(0.0267)	(0.0321)	(0.0387)	(0.0244)	(0.0298)	(0.0257)	(0.0382)	(0.0405)	(0.0289)	(0.0281)
N	5400	5414	5316	3764	5332	4964	5203	3786	3794	3183	3175

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for time, forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

Table 9: Regression Coefficients for three-quarter-ahead Forecasts Only

(a) Bordalo et al. Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	-0.236***	-0.140	-0.206***	-0.461***	0.111	-0.147	-0.285**	-0.302***	0.0998	-0.407***	-0.291***
	(0.0721)	(0.0951)	(0.0629)	(0.0604)	(0.117)	(0.0888)	(0.111)	(0.112)	(0.0873)	(0.0653)	(0.0823)
N	4994	4920	4954	3496	5101	4592	4810	3570	3611	3080	3061

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(b) Coibion-Gorodnichenko Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	0.977***	0.858**	1.090***	0.181	1.306***	1.002***	0.347	0.453	0.504**	-0.113	0.188
	(0.365)	(0.352)	(0.375)	(0.509)	(0.392)	(0.378)	(0.444)	(0.418)	(0.252)	(0.266)	(0.261)
N	194	194	195	193	195	193	194	149	152	152	113

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are robust to arbitrary heteroskedasticity and autocorrelation.

Dummy variables for forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(c) Gemmi-Valchev Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Surprise	0.500***	0.509***	0.545***	0.489***	0.439***	0.467***	0.449***	0.466***	0.507***	0.481***	0.491***
	(0.0291)	(0.0341)	(0.0316)	(0.0286)	(0.0195)	(0.0305)	(0.0318)	(0.0332)	(0.0254)	(0.0228)	(0.0282)
N	5131	5059	5060	3579	5212	4706	4931	3683	3691	3130	3115

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for time, forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

Table 10: Regression Coefficients Using All Sample Periods

(a) Bordalo et al. Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	-0.114**	-0.130**	-0.230***	-0.393***	-0.278***	-0.149	-0.259***	-0.0499	-0.0226	-0.362***	-0.114**
	(0.0573)	(0.0631)	(0.0554)	(0.0660)	(0.0382)	(0.0931)	(0.0698)	(0.123)	(0.0443)	(0.0446)	(0.0439)
N	21925	21904	21547	15391	22362	20309	21017	15366	15516	12871	13192

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(b) Coibion-Gorodnichenko Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Revision	0.152*	0.105	0.772***	0.483***	-0.209***	0.318***	0.201	0.463***	0.162***	-0.147**	0.204***
	(0.0835)	(0.0809)	(0.148)	(0.183)	(0.0369)	(0.121)	(0.138)	(0.107)	(0.0536)	(0.0737)	(0.0429)
N	837	837	837	833	841	837	837	638	646	646	478

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are robust to arbitrary heteroskedasticity and autocorrelation.

Dummy variables for forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

(c) Gemmi-Valchev Specification

	NGDP	RGDP	PGDP	CPROF	UNEMP	INDPROD	HOUSING	CPI	TBILL	BOND	TBOND
Surprise	0.578***	0.631***	0.611***	0.498***	0.783***	0.601***	0.628***	0.589***	0.752***	0.698***	0.704***
	(0.0350)	(0.0277)	(0.0275)	(0.0323)	(0.0782)	(0.0371)	(0.0360)	(0.0408)	(0.0313)	(0.0256)	(0.0267)
N	22308	22297	21916	15641	22643	20629	21354	15652	15680	12976	13354

Significance: *=10%, **=5%; ***=1%. Standard errors in parentheses are two-way clustered in forecaster and time.

Dummy variables for time, forecaster and forecast horizon are controlled. Variables have different sample periods. The longest sample is 1968Q1-2022Q2.

Table 11: Estimationg Using the Long-term CPI Forecasts

	10-Year	5-Year	3-quarter	Current quarter
Surprise	0.0828***	0.129***	0.460***	0.855***
	(0.0183)	(0.0413)	(0.0599)	(0.0310)
N	2672	1613	3496	3602

Significance: *=10%, **=5%; ***=1%.
Standard errors in parentheses are two-way clustered in forecaster and time.
Dummy variables for the forecaster and forecast horizon are controlled.
Variables have different sample periods. The longest sample is 1968Q4-2022Q2.

E.3 Data Generating Process: AR(1)

I use the following steps to set the parameters for each macroeconomic variable y_t. Using the actual realization, I first get the OLS estimates of the AR(1) parameter ρ and σ_ϵ^2. Table 12 reports the parameters.

Table 12: Data Generating Process: AR(1) process

	ρ	σ_ϵ
Nominal Gross Domestic Product	0.89	1.31
Real Gross Domestic Product	0.84	1.32
GDP Chain-Weighted Price Index	0.83	1.31
Corporate Profits After Taxes	0.75	8.03
Industrial Production Index	0.85	2.49
Housing Starts	0.84	11.68
Consumer Price Index	0.75	2.0
AAA Corporate Bond Yield	0.83	0.51
Treasury Bond Rate, 10-year	0.82	0.63

F A Stationary Relationship

In summary, the posterior mean for x_t evolves according to

$$\begin{aligned} x_{i,t|t} &= (I - K_t)\, x^m_{i,t|t} + K_t\, x_t + \bar{\nu}_t + \bar{u}_{i,t} \\ &= (I - K_t)\left((I - \Lambda_t)\, \mu_x + \Lambda_t\, x_{i,t|t-1} + \omega_{i,t}\right) + K_t\, x_t + \bar{\nu}_t + \bar{u}_{i,t} \\ &= (I - K_t)\,(I - \Lambda_t)\, \mu_x + (I - K_t)\, \Lambda_t\, A\, x_{i,t-1|t-1} + K_t\, x_t + \bar{\nu}_t + \bar{\omega}_{i,t} + \bar{u}_{i,t} \end{aligned}$$

where $\bar{\omega}_{i,t} \equiv (I - K)\, \omega_{i,t}$. To ease the notation burden, I define $\Delta_t \equiv (I - K_t)\,(I - \Lambda_t)$, $\hat{c}_t \equiv \Delta_t\, \mu_x$ and $\hat{A}_t \equiv (I - K_t - \Delta_t)\, A$. Then,

$$x_{i,t|t} = \hat{c}_t + \hat{A}_t\, x_{i,t-1|t-1} + K_t\, x_t + \bar{\nu}_t + \bar{\omega}_{i,t} + \bar{u}_{i,t} \tag{53}$$

Stationary relationship As $t \to \infty$, these matrixes converge to a steady state level. Let's denote this as $\bar{K}_t \to \bar{K}$, $K_t \to K$ and $\Lambda_t \to \Lambda$. Then, forecasts for x_t evolve according to

$$x_{i,t|t} = \hat{c} + \hat{A}\, x_{i,t-1|t-1} + K\, x_t + \bar{\nu}_t + \bar{\omega}_{i,t} + \bar{u}_{i,t} \tag{54}$$

where the variance of the noise is

$$\begin{aligned} Var\left[\bar{\nu}_t\right] &= Var\left[\bar{K}\, \nu_{i,t}\right] = \sigma^2_\nu\, \bar{K}\, \bar{K}' \\ Var\left[\bar{u}_{i,t}\right] &= Var\left[\bar{K}\, u_{i,t}\right] = \sigma^2_u\, \bar{K}\, \bar{K}' \\ Var\left[\bar{\omega}_{i,t}\right] &= Var\left[(I - K)\, \omega_{i,t}\right] = (I - K)\, \sigma_{\omega,t}\, (I - K)' = (I - K)\, (I - \Lambda)\, Var\left[x_{i,t|t-1}\right]\, \Lambda'\, (I - K)' \\ &= \Delta\, Var\left[x_{i,t|t-1}\right]\, (I - K - \Delta)' \end{aligned}$$

Statistical properties of $x_{i,t|t}$

Since $x_{i,t|t}$ is a conditional expectation of x_t given available information at time t, we can easily see that following holds.

$$\begin{aligned} Cov\left[x_{i,t|t}, x_t\right] &= Var\left[x_{i,t|t}\right] = \sigma_x - \Sigma_{t|t} \\ Cov\left[x_{i,t|t}, x_{i,t-1|t-1}\right] &= Cov\left[\left(\hat{A} + K\, A\right) x_{i,t-1|t-1}, x_{i,t-1|t-1}\right] = (I - \Delta)\, A\left(\sigma_x - \Sigma_{t|t}\right) \end{aligned}$$

Evolution of the average forecasts for x_t

The average forecasts for x_t evolve according to

$$x_{t|t} = \hat{c} + \hat{A}\, x_{t-1|t-1} + K\, x_t + \bar{\nu}_t \tag{55}$$

Therefore, we can see that

$$x_{i,t|t} - x_{t|t} = \hat{A}\left(x_{i,t-1|t-1} - x_{t-1|t-1}\right) + \bar{\omega}_{i,t} + \bar{u}_{i,t}$$

By iterating backward, we can also see that the difference between $x_{i,t|t}$ and $x_{t|t}$ is the history of noise realizations.

$$x_{i,t|t} - x_{t|t} = \sum_{j=0}^{\infty} \hat{A}^j \left(\bar{\omega}_{i,t-j} + \bar{u}_{i,t-j}\right) \equiv NoiseHistory_{i,t}$$

Therefore, the covariance between $x_{t|t}$ and x_t is the same as the covariance between $x_{i,t|t}$ and x_t. Also, we can express the variance of $x_{t|t}$ as the variance of $x_{i,t|t}$ subtracted by the variance of the history of noises.

$$Var\left[x_{i,t|t}\right] = Var\left[x_{t|t}\right] + Var\left[NoiseHistory_{i,t}\right]$$

where $Var\left[NoiseHistory_{i,t}\right] = Var\left[\sum_{j=0}^{\infty} \hat{A}^j \left(\bar{\omega}_{i,t} + \bar{u}_{i,t}\right) \left(\hat{A}^j\right)'\right]$. Finally, the serial correlation of $x_{t|t}$ is derived.

$$\begin{aligned} Cov\left[x_{t|t}, x_{t-1|t-1}\right] &= \hat{A}\, Var\left[x_{t|t}\right] + K\, A\, Cov\left[x_t, x_{t|t}\right] \\ &= \left(\hat{A} + K\, A\right) Var\left[x_{i,t|t}\right] - \hat{A}\left(I - \hat{A}\right)^{-1} Var\left[\bar{\omega}_{i,t} + \bar{u}_{i,t}\right] \\ &= Cov\left[x_{i,t|t}, x_{i,t-1|t-1}\right] - \hat{A}\left(I - \hat{A}\right)^{-1} Var\left[\bar{\omega}_{i,t} + \bar{u}_{i,t}\right] \end{aligned}$$

Perceived covariance of individual forecast errors and revisions

Given the prior $x_t|m_{i,t} \sim \mathcal{N}\left(x^m_{i,t|t},\ \Sigma^m_{t|t}\right)$, posterior distribution of $x_t|s_{i,t}$ is chosen so that it satisfies $Cov\left[x_t - x_{i,t|t}, x_{i,t|t} - x^m_{i,t|t}\right] = O$. However, we are interested in the covariance between the forecast error and revision observed by an econometrician. To see the difference, it is useful to express the law of motion of $x_{i,t|t}$ as follows.

$$\begin{aligned} x_{i,t|t} &= \left(I - K\right) x^m_{i,t|t} + K\, x_t + \bar{\nu}_t + \bar{u}_{i,t} \\ &= \left(I - K\right) x_{i,t|t-1} + K\, x_t - \left(I - K\right)\left(I - \Lambda\right)\left(x_{i,t|t-1} - \mu_x\right) + \bar{\nu}_t + \bar{u}_{i,t} + \bar{\omega}_{i,t} \end{aligned}$$

From this, we can see that

$$\begin{aligned} Cov\left[x_t - x_{i,t|t}, x_{i,t|t-1}\right] &= Cov\left[E\left[x_t - x_{i,t|t}\middle| s_{i,t-1}\right], E\left[x_{i,t|t-1}\middle| s_{i,t-1}\right]\right] \\ &= Cov\left[E\left[x_t - x_{i,t|t}\middle| s_{i,t-1}\right], x_{i,t|t-1}\right] \\ &= \left(I - K\right)\left(I - \Lambda\right) Var\left[x_{i,t|t-1}\right] \end{aligned}$$

Therefore,

$$\begin{aligned}
Cov\left[x_t - x_{i,t|t}, x_{i,t|t} - x_{i,t|t-1}\right] &= Cov\left[x_t - x_{i,t|t}, \left(x_{i,t|t} - x^m_{i,t|t}\right) - \left(x_{i,t|t-1} - x^m_{i,t|t}\right)\right] \\
&= -Cov\left[x_t - x_{i,t|t}, x_{i,t|t-1} - x^m_{i,t|t}\right] \\
&= -Cov\left[x_t - x_{i,t|t}, x_{i,t|t-1}\right]\,(I-\Lambda)' - Cov\left[x_{i,t|t}, \omega_{i,t}\right] \\
&= -(I-K)(I-\Lambda)\,Var\left[x_{i,t|t-1}\right](I-\Lambda)' - (I-K)\,Var\left[\omega_{i,t}\right] \\
&= -(I-K)(I-\Lambda)\,Var\left[x_{i,t|t-1}\right] = -(I-K)\left(\sigma^m_{t|t} - \sigma_{t|t-1}\right)
\end{aligned}$$

the last equality follows from $Var\left[\omega_{i,t}\right] = (I-\Lambda)\,Var\left[x_{i,t|t-1}\right]\Lambda'$. If memory is perfect, we can confirm that forecast error would not be predicted by forecast revision.

Perceived covariance of average forecast errors and revisions

The statistical properties of the average forecast $x_{t|t}$ are determined from the following law of motion.

$$x_{t|t} = (I-K)\,x_{t|t-1} + K\,x_t - (I-K)(I-\Lambda)\left(x_{t|t-1} - \mu_x\right) + \bar{\nu}_t$$

Rearranging terms yields

$$K\left(x_t - x_{t|t}\right) = (I-K)\left\{\left(x_{t|t} - x_{t|t-1}\right) + (I-\Lambda)\left(x_{t|t-1} - \mu_x\right)\right\} - \bar{\nu}_t$$

If K is invertible,

$$\begin{aligned}
&Cov\left[x_t - x_{t|t}, x_{t|t} - x_{t|t-1}\right] \\
&= K^{-1}(I-K)\left\{Var\left[x_{t|t} - x_{t|t-1}\right] + (I-\Lambda)\,Cov\left[x_{t|t-1}, x_{t|t} - x_{t|t-1}\right]\right\} - K^{-1}V_{noise}
\end{aligned}$$

Correct covariance of forecast errors and revisions

While DM is uncertain about the value of μ, in reality, μ is a fixed parameter. Therefore, the OLS regression of forecast error on revision will asymptotically converge to the covariance of forecast error and revision arising from a fixed parameter. I show here how such statistics differ from the one derived above (where the covariances are averaged across all possible values of μ according to DM's prior about μ). For individual forecasts,

$$\begin{aligned}
&Cov\left[x_t - x_{i,t|t}, x_{i,t|t} - x_{i,t|t-1}\middle|\,\mu\right] \\
&= Cov\left[x_t - x_{i,t|t}, x_{i,t|t} - x_{i,t|t-1}\right] - Cov\left[E\left[x_t - x_{i,t|t}\middle|\,\mu\right], E\left[x_{i,t|t} - x_{i,t|t-1}\middle|\,\mu\right]\right]
\end{aligned}$$

Likewise, for average forecasts,

$$\begin{aligned}
&Cov\left[x_t - x_{t|t}, x_{t|t} - x_{t|t-1}\right|\mu] \\
&= Cov\left[x_t - x_{t|t}, x_{t|t} - x_{t|t-1}\right] - Cov\left[E\left[x_t - x_{t|t}\right|\mu\right], E\left[x_{t|t} - x_{t|t-1}\right|\mu]]
\end{aligned}$$

Note that the subtracted terms in the above two cases are the same since it must be that $E\left[x_{i,t|t}\right|\mu] = E\left[x_{t|t}\right|\mu]$ at all t. This term is non-zero because forecasts are biased even in the long run, as DM fails to learn the correct level of μ. Using the fact that forecasts for x_t are stationary, that is $E\left[x_{i,t|t}\right|\mu] = E\left[x_{i,t-1|t-1}\right|\mu]$, we have

$$E\left[x_{i,t|t}\right|\mu] = \left(I - \hat{A}\right)^{-1}(\Delta\,\mu_x + K\,E\left[x\right|\mu])$$

Since we can express $E\left[x_{i,t|t}\right|\mu] = cons + D\,E\left[x_t\right|\mu]$, where $D \equiv \left(I - \hat{A}\right)^{-1} K$, the correction term can be derived as

$$Cov\left[E\left[x_t - x_{i,t|t}\right|\mu\right], E\left[x_{i,t|t} - x_{i,t|t-1}\right|\mu]] = (I - D)\,Var\left[E\left[x_t\right|\mu]\right]\,D'\,(I - A)'$$

Correct variance of revisions

The individual and average forecast revision variances are derived and can be computed using the previously derived stationary relationship.

$$\begin{aligned}
Var\left[x_{i,t|t} - x_{i,t|t-1}\right] &= Var\left[x_{i,t|t}\right] + Var\left[x_{i,t|t-1}\right] - Cov\left[x_{i,t|t}, x_{i,t|t-1}\right] - Cov\left[x_{i,t|t-1}, x_{i,t|t}\right] \\
Var\left[x_{t|t} - x_{t|t-1}\right] &= Var\left[x_{t|t}\right] + Var\left[x_{t|t-1}\right] - Cov\left[x_{t|t}, x_{t|t-1}\right] - Cov\left[x_{t|t-1}, x_{t|t}\right]
\end{aligned}$$

The correct variance of the forecast revision (conditional on a fixed μ) is

$$\begin{aligned}
Var\left[x_{i,t|t} - x_{i,t|t-1}\right|\mu] &= Var\left[x_{i,t|t} - x_{i,t|t-1}\right] - Var\left[E\left[x_{i,t|t} - x_{i,t|t-1}\right|\mu]\right] \\
Var\left[x_{t|t} - x_{t|t-1}\right|\mu] &= Var\left[x_{t|t} - x_{t|t-1}\right] - Var\left[E\left[x_{t|t} - x_{t|t-1}\right|\mu]\right]
\end{aligned}$$

The correction term is the same for average and individual forecasts and is derived as follows.

$$Var\left[E\left[x_{i,t|t} - x_{i,t|t-1}\right|\mu]\right] = Var\left[(I - A)\,D\,E\left[x_t\right|\mu]\right] = (I - A)\,D\,Var\left[E\left[x_t\right|\mu]\right]\,D'\,(I - A)'$$

Gemmi-Valchev Proposal

From (54) and (55), forecast revisions are expressed as follows.

$$\begin{aligned}
x_{i,t|t} - x_{i,t|t-1} &= \hat{c} + K\left(x_t - x_{i,t|t-1}\right) - \Delta\,x_{i,t|t-1} + \bar{\nu}_t + \bar{\omega}_{i,t} + \bar{u}_{i,t} \\
x_{t|t} - x_{t|t-1} &= \hat{c} + K\left(x_t - x_{t|t-1}\right) - \Delta\,x_{t|t-1} + \bar{\nu}_t
\end{aligned}$$

Then, the difference between individual and consensus forecast revisions (in other words, de-meaned individual forecast revisions) can be derived as

$$\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right)=(K+\Delta)\left(x_{t|t-1}-x_{i,t|t-1}\right)+\bar{\omega}_{i,t}+\bar{u}_{i,t}$$

The authors propose to estimate the covariance between the de-meaned forecast revisions and the difference between consensus and individual forecasts from the previous period.

$$Cov\left[\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right),x_{t|t-1}-x_{i,t|t-1}\right]=(K+\Delta)\,Var\left[x_{t|t-1}-x_{i,t|t-1}\right]$$

Note that this regression coefficient is well-defined only if individual forecasts deviate from the consensus forecasts (that is, when $\phi_n > 0$ or $\phi_m > 0$).

Furthermore, we can see that the perceived covariance (based on DM's prior about μ) is the same as the correct covariance (given a fixed μ), unlike the other covariances I derived earlier.

$$\begin{aligned}&Cov\left[\left.\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right),x_{t|t-1}-x_{i,t|t-1}\right|\mu\right]\\&=Cov\left[\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right),x_{t|t-1}-x_{i,t|t-1}\right]\end{aligned}$$

This is because the correction term cancels out by de-meaning. That is,

$$\begin{aligned}E\left[\left.\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right)\right|\mu\right]&=0\\E\left[\left.x_{t|t-1}-x_{i,t|t-1}\right|\mu\right]&=0\end{aligned}$$

This result follows from $E\left[\left.x_{i,t|t}\right|\mu\right]=E\left[\left.x_{t|t}\right|\mu\right]$ and $E\left[\left.x_{i,t|t-1}\right|\mu\right]=E\left[\left.x_{t|t-1}\right|\mu\right]$.

To compute the regression coefficient, it remains to derive an expression for $Var\left[x_{t|t-1}-x_{i,t|t-1}\right]$. Note that

$$x_{t|t}-x_{i,t|t}=\hat{A}\left(x_{t-1|t-1}-x_{i,t-1|t-1}\right)-\left(\bar{\omega}_{i,t}+\bar{u}_{i,t}\right)$$

Therefore, $Var\left[x_{t|t}-x_{i,t|t}\right]$ satisfies the following fixed-point relation.

$$Var\left[x_{t|t}-x_{i,t|t}\right]=\hat{A}Var\left[x_{t|t}-x_{i,t|t}\right]\hat{A}'+Var\left[\bar{\omega}_{i,t}+\bar{u}_{i,t}\right]$$

We can then derive

$$Var\left[x_{t|t-1}-x_{i,t|t-1}\right]=A\,Var\left[x_{t|t}-x_{i,t|t}\right]\,A'$$

Finally, for a given c',

$$\beta=\frac{c'\,(K+\Delta)\,Var\left[x_{t|t-1}-x_{i,t|t-1}\right]\,c}{c'\,Var\left[x_{t|t-1}-x_{i,t|t-1}\right]\,c}$$

Furthermore, the variance of the de-meaned forecast revisions is derived as

$$Var\left[\left(x_{i,t|t}-x_{i,t|t-1}\right)-\left(x_{t|t}-x_{t|t-1}\right)\right]=(K+\Delta)\,Var\left[x_{t|t-1}-x_{i,t|t-1}\right](K+\Delta)'+Var\left[\bar{\omega}_{i,t}+\bar{u}_{i,t}\right]$$

and it must be that $Var\left[\left(x_{i,t|t} - x_{i,t|t-1}\right) - \left(x_{t|t} - x_{t|t-1}\right)\middle|\, \mu\right] = Var\left[\left(x_{i,t|t} - x_{i,t|t-1}\right) - \left(x_{t|t} - x_{t|t-1}\right)\right]$.

G Monetary Model

I describe a textbook model below, but more details can be found in Gali (2008, Chapter 3).

G.1 Household Problem

A representative, infinitely-lived household maximizes the lifetime utility from consumption and labor.

$$E_0 \sum_{t=0}^{\beta} \left[\frac{C_t^{1-\sigma}}{1-\sigma} - \frac{N_t^{1+\varphi}}{1+\varphi} \right]$$

where C_t is the quantity of the basket of goods consumed at time t, and N_t is the number of hours worked. The consumption/savings and labor-supply decisions are subject to the budget constraint that should be met every period.

$$P_t\, C_t + Q_t\, B_t \leq B_{t-1} + W_t\, N_t + T_t$$

where P_t is the aggregate price index, B_t is the one-period bond and Q_t its price, W_t is the nominal hourly wage, and finally T_t is a lump-sum income. The household should also be solvent after all, which is captured by the condition that $\lim_{T\to\infty} Et\, B_t \geq 0$.

The first order conditions and their Taylor expansion around the zero-inflation steady state imply

$$w_t - p_t = \sigma\, c_t + \varphi\, n_t \tag{56}$$

$$c_t = E_t\, c_{t+1} - \frac{1}{\sigma}\,(-q_t - E_t\, \pi_{t+1} + log\, \beta) \tag{57}$$

where the lowercase denotes the log of the variable denoted in uppercase.

G.2 Firm Problem

A continuum of firms indexed by $i \in [0,1]$ produces a differentiated goods. The production function is described as

$$Y_t(i) = A_t\, N_t(i)$$

where A_t is the level of production technology, assumed to be common to all firms and evolve exogenously over time.

Each firm reconsiders its price with probability $1-\alpha$, independent of when its price is readjusted in the past. Thus, at any period, a mass of $1-\alpha$ firms resets their prices and the remaining mass of

α firms keep their old prices. The aggregate price index is then formed according to

$$P_t = \left[\alpha \left(P_{t-1}\right)^{1-\eta} + (1-\eta)\left(\int P^*_{i,t}\, di\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}$$

G.3 Optimal Price Setting

Suppose firm i chooses the price $P^*_{i,t}$ in period t. This price maximizes the current market value of the profits if the firm cannot reoptimize the price forever.

$$\max_{P_{i,t}} \quad E_{i,t}\left[\sum_{h=0}^{\infty} \alpha^h\, Q_{t,t+h}\left(P_{i,t}\, Y_{i,t+h|t} - \Psi_{t+h}\left(Y_{i,t+h|t}\right)\right)\right]$$

where α the probability of not resetting prices, $Q_{t,t+h}$ is the stochastic discount factor for evaluating the future nominal payoffs generated at $t+h$, $Y_{i,t+h|t}$ is the output demanded in period $t+h$ if the price remains the one chosen at time t, and Ψ_{t+h} is the (nominal) cost function at time $t+h$. Firm i takes into account that the demand $Y_{i,t+h|t}$ is given as

$$Y_{i,t+h|t} = \left(\frac{P_{i,t}}{P_{t+h}}\right)^{\eta} C_{t+h}$$

where θ is the elasticity of substitution among goods, P_{t+k} is the aggregate price at time $t+h$ and C_{t+h} is the aggregate consumption at time $t+h$.

The first-order condition implies that

$$E_{i,t}\left[\sum_{h=0}^{\infty} \alpha^h\, Q_{t,t+h} Y_{i,t+h|t}\left(P^*_{i,t} - \mathcal{M}\psi_{t+h}\right)\right] = 0$$

where $\mathcal{M} \equiv \frac{\eta}{\eta-1}$ and ψ_{t+h} is the nominal marginal cost at $t+h$. Dividing by P_{t-1} and letting $\Pi_{t,t+h} \equiv \frac{P_{t+h}}{P_t}$, we can rewrite the first order condition as

$$E_{i,t}\left[\sum_{h=0}^{\infty} \alpha^h\, Q_{t,t+h} Y_{i,t+h|t}\left(\frac{P^*_{i,t}}{P_{t-1}} - \mathcal{M} MC_{t+h}\, \Pi_{t,t+h}\right)\right] = 0$$

First-order Taylor expansion around the zero-inflation steady state implies that

$$\begin{aligned} p^*_{i,t} - p_{t-1} &= E_{i,t}\left[(1-\alpha\beta)\sum_{h=0}^{\infty} (\alpha\beta)^h \left((mc_{t+h} - mc) + (p_{t+h} - p_{t-1})\right)\right] \\ &= E_{i,t}\left[\sum_{h=0}^{\infty} (\alpha\beta)^h \left\{(1-\alpha\beta)\left(mc_{t+h} - mc\right) + \pi_{t+h}\right\}\right] \end{aligned}$$

where mc is the steady state value of mc_{t+h}. From this expression, we can see that the optimal reset price $p^*_{i,t}$ equals mc over a weighted average of the current and expected nominal marginal costs.

Note that the marginal cost at $t+h$ does not depend on the quantity firm i supplies. This is because the marginal product of labor does not depend on quantity, as $mpn_t = a_t$. Thus,

$$mc_{t+h} = w_{t+h} - p_{t+h} - mpn_{t+h} = w_{t+h} - p_{t+h} - a_{t+h}$$

G.4 Equilibrium

Since market clears for all i goods, it follows that

$$C_t = Y_t$$

which implies $c_t = y_t$. And the labor market clears, requiring

$$N_t = \int N_t(i)\, di$$

which can be shown to imply $n_t = y_t - a_t$ in the first order approximation. Thus, using the household's optimality condition,

$$w_t - p_t = (\sigma + \varphi)\, y_t - \varphi\, a_t$$

Denoting y_t^n as the efficient level of output, we can show that $y_t^n = \frac{1+\varphi}{\sigma+\varphi}\, a_t$. I define the output gap as

$$x_t = y_t - y_t^n$$

Thus, the marginal costs are derived as

$$mc_{t+h} = (\sigma + \varphi)\, x_t$$

G.5 Firms' Macroeconomic Expectations

Substituting (22), we can see that inflation is determined as

$$\pi_t = (1-\alpha)\left(\hat{z}_t + \frac{\alpha\beta}{1-\alpha\beta}\,\hat{\mu}_t\right)$$

Substituting (23) and (24), we get

$$\pi_t = (1-\alpha)\left\{(\kappa + \kappa_\mu)\, z_t + \left(\lambda(1-\kappa) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)\right)\hat{\mu}_{t-1}\right\}$$

Defining $\hat{\kappa} = \kappa + \kappa_\mu$ and $\hat{b} = \lambda(1-\kappa) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)$, we can describe the above expression as

$$\pi_t = (1-\alpha)\left\{\hat{\kappa}\, z_t + \hat{b}\,\hat{\mu}_{t-1}\right\} \tag{58}$$

G.6 Inflation Determination

We can solve for the equilibrium inflation process using a guess-and-verify approach. The equation (21) states that z_t is determined by π_t and e_t, and the equation (58) states that π_t is determined by z_t and $\hat{\mu}_{t-1}$. Thus, it is straightforward to see that two state variables, e_t and $\hat{\mu}_{t-1}$, determine inflation, and the relationship is linear. We guess the following inflation process.

$$\pi_t = \varphi_e \, e_t + \varphi_\mu \, \hat{\mu}_{t-1} \tag{59}$$

Combining (21), (58), and (59), we can find the coefficients φ_e and φ_μ that verify our initial guess. They are derived as below.

$$\varphi_e = \frac{\delta}{1 + \delta \, \sigma \, s + \frac{1}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}}$$

$$\varphi_\mu = \frac{1}{1 + \delta \, \sigma \, s + \frac{1}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}} \frac{1-\alpha}{\alpha} \frac{\hat{b}}{\hat{\kappa}}$$

where $\delta \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$, $\hat{\kappa} = \kappa + \kappa_\mu$, and $\hat{b} = \lambda(1-\kappa) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)$. When exploring alternative monetary policies, I consider values of s such as

$$s = s^* \cdot \frac{\theta}{1-\theta}$$

in which case, we could express the coefficients φ_e and φ_μ as

$$\varphi_e = \frac{1-\theta}{1 - \theta + \theta \, \hat{\delta} \, s^* + \frac{1-\theta}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}} \, \delta$$

$$\varphi_\mu = \frac{1-\theta}{1 - \theta + \theta \, \hat{\delta} \, s^* + \frac{1-\theta}{\alpha} \frac{1-\hat{\kappa}}{\hat{\kappa}}} \frac{1-\alpha}{\alpha} \frac{\hat{b}}{\hat{\kappa}}$$

This expression makes it clear that a complete inflation stabilization ($\theta = 1$) is supported by $\varphi_e = \varphi_m = 0$.

G.7 Variability of Inflation

From (59), we can see that the variability of inflation is derived as

$$Var\left[\pi_t\right] = \varphi_e^2 \, Var\left[e_t\right] + \varphi_\mu^2 \, Var\left[\hat{\mu}_{t-1}\right]$$

Therefore, it remains to derive the variability of $\hat{\mu}_t$. First, note that from (21), z_t is also determined by two state variables.

$$z_t = \varpi_e \, e_t + \varpi_\mu \, \hat{\mu}_{t-1}$$

where ϖ_e and ϖ_μ are defined as

$$\varpi_e = \frac{\alpha}{1-\alpha}\left(\left(\frac{1}{\alpha} - \frac{1-\theta+\theta\hat{\delta}}{1-\theta}\right)\varphi_e + \delta\right)$$
$$\varpi_\mu = \frac{\alpha}{1-\alpha}\left(\frac{1}{\alpha} - \frac{1-\theta+\theta\hat{\delta}}{1-\theta}\right)\varphi_\mu$$

Using this expression, we can then describe the law of motion of $\hat{\mu}_t$ as

$$\hat{\mu}_t = \underbrace{\left(\lambda\left(1-\kappa_\mu\right) + \kappa_\mu\,\varpi_\mu\right)}_{\equiv\rho_\mu}\hat{\mu}_{t-1} + \kappa_\mu\,\varpi_e\,e_t$$

From this, we can see that

$$Var\left[\hat{\mu}_t\right] = \frac{\left(\kappa_\mu\,\varpi_e\right)^2}{1-\rho_\mu^2}\,Var\left[e_t\right]$$

Therefore, the variability of inflation is derived as

$$Var\left[\pi_t\right] = \left(\varphi_e^2 + \frac{\left(\kappa_\mu\,\varpi_e\right)^2}{1-\rho_\mu^2}\right)Var\left[e_t\right]$$