FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Macroeconomic Expectations and Cognitive Noise

Yeji Sung
Federal Reserve Bank of San Francisco

June 2024

Working Paper 2024-19

https://doi.org/10.24148/wp2024-19

Suggested citation:

Sung, Yeji. 2024. "Macroeconomic Expectations and Cognitive Noise." Federal Reserve Bank of San Francisco Working Paper 2024-19. https://doi.org/10.24148/wp2024-19

Macroeconomic Expectations and Cognitive Noise*

Yeji Sung†

Federal Reserve Bank of San Francisco

June 14, 2024

Abstract

This paper examines forecast biases through cognitive noise, moving beyond the conventional view that frictions emerge solely from using external data. By extending Sims's (2003) imperfect attention model to include imperfect memory, I propose a framework where cognitive constraints impact both external and internal information use. This innovation reveals horizon-dependent forecast sensitivity: short-term forecasts adjust sluggishly while long-term forecasts may overreact. I explore the macroeconomic impact of this behavior, showing how long-term expectations, heavily influenced by current economic conditions, heighten inflation volatility. Moreover, structural estimation indicates that neglecting imperfect memory critically underestimates the informational challenges forecasters encounter.

JEL codes: D84, E32, E71, G41
Keywords: Information Frictions, Rational Inattention, Business Cycle Fluctuations

*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, the Federal Reserve Bank of St. Louis, the Federal Reserve Bank of San Francisco, and many other seminars/conferences. 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.

†Email:yeji.sung@sf.frb.org.

Introduction

The persistent presence of bias in economic forecasts raises a fundamental question: What limits our ability to generate unbiased predictions? Despite extensive research, including analyses of survey forecasts and controlled experiments, a comprehensive yet tractable model of expectations formation remains elusive. This gap poses a particular challenge for macroeconomic analysis, where dynamic model predictions rely heavily on assumptions about how economic agents form expectations. An empirically grounded model of this process would enhance our ability to evaluate the validity of macroeconomic models.

Traditionally, macroeconomic forecast biases have been attributed to information frictions – imperfect knowledge of the economic environment. Such models posit that limitations in accessing and processing relevant data lead to gradual updates in expectations, a pattern consistent with some features observed in macroeconomic survey data (e.g., Coibion and Gorodnichenko (2012, 2015)). However, this framework provides an incomplete explanation, as forecasts often exhibit substantial overreactions to new information. This pattern challenges the predictions of models relying solely on traditional information frictions, prompting researchers to explore additional separate mechanisms (e.g., Bordalo et al. (2020), Angeletos et al. (2021), and Gemmi and Valchev (2023)).

In this paper, I argue that a more coherent explanation for forecast biases lies in expanding the concept of information frictions to include limitations in processing both external and internal information sources. To this end, I introduce a new model where imperfect attention and memory hinder the seamless utilization of both types of information. This framework introduces a new mechanism: the inability to seamlessly integrate past knowledge into new forecasts. Consequently, prior beliefs play a diminished role in resolving economic uncertainty, leading forecasters to overemphasize recent news. This model explains the pervasive patterns in survey forecasts of macroeconomic and financial variables documented in Coibion and Gorodnichenko (2015) and Bordalo et al. (2020).

A key prediction of the model is that forecasts may exhibit horizon-dependent sensitivities to incoming data. For near-term projections, forecasters tend to revise slowly and in a muted fashion, underreacting to recent information. Conversely, long-term forecasts are prone to overreacting relative to the predictions of fully frictionless models without any informational constraints. This prediction aligns well with biases observed in survey forecasts (e.g., Bordalo et al. (2023), d'Arienzo (2020), Wang (2021), and Shiller and Thompson (2022)). The macroeconomic implications of this finding are demonstrated through an analysis of inflation dynamics. Even with a fixed inflation target, the model shows that firms' long-term expectations about the cost conditions shift along with the current economic environment, leading to more volatile inflation dynamics. This makes it harder for the central bank to balance between stabilizing inflation and the output gap.

Finally, I present a structural estimation of the underlying information frictions. This estimation demonstrates how such frictions contribute to the biases observed in professional forecasts. It further suggests that conventional models likely underestimate the severity of information frictions by assuming perfect memory. My estimates suggest the the extent of frictions quantified by mutual information are 50 percent larger, once we account for frictions in processing both external and internal information. Thus, accounting for memory limitations reveals that informational constraints are more binding than previously thought.

One might wonder why even professional forecasters would grapple with information frictions. After all, accessing official statistics doesn't seem overly costly, suggesting there should be little friction in gathering these data, nor any significant memory challenges. However, this view overlooks the nuanced art of forecasting; forecasters track and combine a vast amount of dynamic data sources, both traditional and nontraditional, to produce forecasts. Some of these sources provide timely insights on emerging economic issues not yet reflected (or perhaps never captured) in official statistics. This reality poses two key problems: first, processing all such real-time data alongside conventional sources is cognitively demanding, as highlighted by traditional information friction models. Second, forecasters likely rely on memory to recall and contextualize some past information, despite the ability to access it again. In other words, drawing upon one's accumulated experience remains vital given the cognitive costs of continuously processing the full universe of external data in real time.

To clarify the distinct friction sources, I distinguish between *external data* that forecasters can access (e.g., data releases or FOMC press conferences) and *internal memory* that stores accumulated knowledge from prior experience. Consistent with conventional models, I posit that forecasters face constraints in processing the full breadth of external data flows. However, I depart from tradition by also accounting for frictions in recalling and integrating one's knowledge repository. Specifically, forecasts get produced not with perfect recall but with imperfect memory; preexisting knowledge is not seamlessly integrated into new forecasts. Rather than assuming ever-expanding, perfectly nested information sets, I conceptualize memory as a stochastically accessed collection of past knowledge.

Given the information-processing constraint, I propose a model accommodating the flexible specification of attention and memory systems. Rather than imposing an ad hoc structure, this model posits that the cognitive process adapts to prioritize the most task-relevant information for forecasting the latent variable. Allocation of attention resources and the strength of specific memory recall are flexible within this framework. Crucially, the model maintains parsimony with two core parameters encapsulating the severity of information-processing constraints: one for attention and the other for memory. These constraints shape the structure of attentional and memory noise, reflecting the premise of adaptive yet inherently noisy information-processing mechanisms. This framework also demonstrates scalability. Under specified assumptions, I derive optimal information processes applicable to any linear state-space system (regardless of dimension), detailed in the appendix.

I demonstrate that forecast biases arising from cognitive noise align with the biases observed in survey forecasts. Notably, the model explains deviations from full-information rational expectations documented in Coibion and Gorodnichenko (2015) and Bordalo et al. (2020). These studies offer contrasting perspectives on the role of information frictions in explaining forecast biases. Coibion and Gorodnichenko (2015) analyze forecasts across macroeconomic and financial variables, finding a pattern of underrevision in consensus forecasts. This supports their view that limitations in accessing and processing external data drive forecast bias. However, Bordalo et al. (2020) challenge this conclusion, identifying biases within individual forecasts from the same survey data that traditional information friction models cannot explain. They find that a forecaster's recent revisions predict their errors; in particular, an upward revision is associated with forecasts that turn out to exceed realizations. This pattern suggests that individual forecasters are more responsive to new information than traditional models would predict.

My model proposes memory noise as a crucial element in reconciling these empirical findings. While attention noise offers a compelling explanation for underrevision in consensus forecasts (stemming from the imperfect incorporation of external data), it falls short in addressing patterns observed at the individual level. If attention noise was the only friction, the errors in individual forecasts should not be predictable by information that individual forecasters ought to have, such as their own forecast revisions. This is the prediction of traditional information friction models based on the assumption that individuals efficiently use available information, both current and past. In comparison, memory noise directly influences the degree to which forecasters adjust their beliefs in response to new information. When prior knowledge becomes less reliable due to forgetfulness, it carries less weight in resolving uncertainty about the economic state. In turn, forecasters prioritize newly acquired information. Consequently, memory noise induces overreliance on recent information when updating forecasts, an effect not predicted by models focusing solely on attention noise.

Importantly, the proposed framework predicts that forecast sensitivity to incoming data varies with the forecast horizon. Specifically, near-term forecasts tend to exhibit underreaction, while long-term forecasts are more likely to display overreaction (compared to the benchmark free of cognitive noise). Noisy news alone generates underreaction, and noisy memory alone generates overreaction. Given the presence of both noises, the relative dominance of the two thus matters. When discussing the different reaction sensitivity over forecast horizons, it is useful to clarify the benchmark Kalman gains (without any cognitive noise). Because external information is more directly informative about the near-term state, the benchmark gains are high. Thus, it requires larger degrees of noisy memory to push the Kalman gains above this higher benchmark level. The opposite is true for long-term forecasts; with smaller long-term benchmark gains, relatively smaller levels of noisy memory are sufficient to cause overreaction. Consequently, a model with both attention and memory noise can predict a spectrum of forecast responses, ranging from underreaction in the near term to overreaction for longer horizons. This prediction is consistent with the findings in Wang (2021) and d'Arienzo (2020) that forecast for longer horizon interest rates displays more sensitivity to incoming data.

I discuss the macroeconomic implications of the proposed cognitive noise by examining their impact on the inflation process. Using a standard New Keynesian framework, where firms set prices according to their macroeconomic expectations, I demonstrate how alternative expectation formation mechanisms can result in inflation dynamics that are qualitatively different. When price-setters are subject to the cognitive constraints introduced in this paper, inflation stabilization becomes significantly more challenging than within conventional information friction models. This stems from the lack of anchoring in the long-run economy. Since firms lack perfect awareness about the long-run cost conditions, their beliefs persistently fluctuate. This additional volatility propagates through price-setting decisions, resulting in a more volatile inflation process. Such increases in volatility challenge the central bank's policy trade-off between inflation and output stabilization.

Finally, I employ the proposed model to estimate the severity of cognitive constraint present in survey forecasts. In particular, I analyze professional forecasters' projections for the gross domestic output measures to infer the extent of attention and memory constraints. My estimate of the attention constraint is roughly 50 percent larger than that of Coibion and Gorodnichenko (2015), who implicitly assumed perfect memory in their analysis of information frictions. Why does assuming perfect memory lead to an underestimation of underlying attention constraints? This occurs because more aggressive forecast revisions, driven by imperfect memory, are misinterpreted as stemming from laxer attention constraints. Using the estimated model, I demonstrate that cognitive noise alone can account for a substantial portion of the observed variation in forecasts and forecast revisions.

My proposed model of expectations formation offers a parsimonious framework to explain the puzzling features of survey forecasts. A single type of information friction – the finite capacity to process vast amounts of data – prevents economic agents from generating forecasts consistent with full information rational expectations (FIRE). This stands in contrast to past literature, where non-Bayesian assumptions, in addition to information frictions, are often invoked to explain forecast biases. For example, researchers have proposed representative heuristics (Bordalo et al. (2020)), model mis-specification (Angeletos et al. (2021)), and incentives to stand out (Gemmi and Valchev (2023)) as an additional mechanism, as explanations for patterns in Coibion and Gorodnichenko (2015) and Bordalo et al. (2020). While insightful, these approaches do not provide reasons why such mechanisms should exist alongside information frictions. My model demonstrates that cognitive noise alone can address these patterns. Importantly, it also explains why under/overreactions may depend on the forecast horizon – a feature that previous proposals fail to capture.

This paper also contributes to the understanding of long-run expectations, a topic of growing interest due to its role in recent inflation trends (e.g., Carvalho et al. (2023) and Hazell et al. (2022)). My model predicts that, even with a fixed central bank inflation target, long-run expectations will fluctuate in response to current economic environments. This stems from memory frictions: discounting of past data hinders accurate learning of the long-run state of the economy. While structural shifts in the economy are a valid reason to discount past data (e.g., Crump et al. (2023) and Farmer et al. (2024)), this cannot fully explain experimental findings that demonstrate overreaction to recent data in environments with a known, constant mean (Afrouzi et al. (2023)). Importantly, clarifying the possible source of bias in expectations formation has direct implications for monetary policy (for example, see Orphanides and Williams (2006) and Gáti (2023)). If long-term expectations fluctuate solely due to changing inflation targets, the concern for policy may be limited. However, my model suggests that persistent policy challenges arise when bouts of inflation can distort long-term expectations away from the clearly communicated set target.

This paper proceeds as follows. Section 1 introduces a model of expectations formation incorporating cognitive noise. Section 2 examines how this model's predictions align with forecast-revision patterns observed in macroeconomic survey data. Section 3 analyzes the model's implications for the term structure of expectations. Section 4 demonstrates the importance of these findings by comparing how different expectations-formation models influence the stochastic properties of inflation. Section 5 presents structural estimation results, and Section 6 concludes.

1 Model: Cognitive Noise in Expectation Formation

This section proposes a model of how people form expectations, considering the limitations of our brains in processing information. The model posits that individuals form their expectations based on available information, but this process is subject to random errors or distortions due to limited attention and memory constraints. As a result, individuals' expectations may not perfectly align with the true underlying probabilities.

1.1 The Forecasting Problem

A macroeconomic variable of interest is y_t, whose stochastic process is composed of two processes, z_t and η_t, as described by the following data-generating process.

$$y_t = z_t + \eta_t \tag{1.1}$$

where z_t captures the persistent business cycle variations, and η_t is the transitory component of y_t. I further suppose that z_t is described as an auto-regressive process

$$z_t = (1-\rho)\,\mu + \rho\, z_{t-1} + \epsilon_t \tag{1.2}$$

where μ is the long-run mean of z_t, and ρ is its serial correlation (with $|\rho| < 1$). The innovation ϵ_t is assumed to be drawn from a Gaussian distribution $\mathcal{N}\left(0,\, \sigma_\epsilon^2\right)$. The variations in η_t are assumed to be drawn from a Gaussian distribution $\mathcal{N}\left(0,\, \sigma_\eta^2\right)$.

I suppose that forecasters have a correct understanding of the parameters describing the data-generating process of y_t (and those of z_t and η_t). However, they do not separately observe realizations of z_t and η_t. Thus, they infer the underlying economic state from available information, which is detailed in the following segment. The task of the forecasters is to make predictions for the probable future values of y_t at each time period. The following expected quadratic loss function captures the lifetime losses from inaccurate projections.

$$E\left[\sum_{t=0}^{\infty} \beta^t \sum_{h=1}^{H} \left(y_{t+h} - F_{i,t}\, y_{t+h}\right)^2\right] \tag{1.3}$$

where $F_{i,t}\, y_{t+h}$ denotes forecaster i's projection of y_{t+h} based on information available at time t, and H is the longest horizon that forecasters make projections for. By (1.1) and (1.2), a forecaster i's projection of y_{t+h} will be

$$F_{i,t}\, y_{t+h} = F_{i,t}\, z_{t+h} = \left(1-\rho^h\right)\mu + \rho^h\, F_{i,t}\, z_t. \tag{1.4}$$

Thus, these forecasts are based on one's projections of the evolution of the hidden state z_t.

1.2 The Cognitive Process

Accurate economic forecasts hinge on the decision-maker's (DM) ability to understand the underlying state z_t. In the cognitive model I propose, the process of acquiring and processing information is inherently imperfect; our cognitive processes introduce random noise, creating a gap between the true state and the DM's understanding of it. This section outlines the source of this "cognitive noise" and propose a mathematical framework to model its impact on forecasting.

The cognitive constraint. I propose that DM's forecasts are not derived directly from the complete set of available information. Instead, they are shaped by a "mental representation" constructed within the DM's mind, a concept well-established in psychology and cognitive science literature (e.g., Paivio (1990)). This representation is inherently less precise than original data, highlighting the complexity faced by forecasters in distilling insights from diverse information sources.

To measure the accuracy of mental representations, I follow Sims (2003) and employ coding theory. A more accurate representation reduces uncertainty about the original information, and this reduction is quantified by the mutual information ($\mathcal{I}$) between the two. I posit that our cognitive processes have limited precision, imposing a constraint on the achievable value of $\mathcal{I}$.

Available information. Information available to forecasters can be divided into two categories. The first is observable data (such as data releases, news articles, etc) which provides updates on the current economic state (z_t). This data is assumed to be accessible to all forecasters. The second category is internal knowledge, consisting of an individual's understanding built from past forecasting experiences. Internal knowledge may vary between forecasters, even when they access the same observable data, as they may have different perspectives. The core distinction here is whether information is external (observable data) or internal (accumulated knowledge) to the DM's mind. Mathematical formulation is provided in the section.

The dual processes. I further propose a cognitive process with two distinct subsystems. The first, which I term the attention system, focuses on external information. It regulates the allocation of mental resources towards relevant external stimuli, aligning with the "rational inattention" literature (Sims (2003), Maćkowiak and Wiederholt (2009), Kacperczyk et al. (2016), Miao et al. (2022), Afrouzi and Yang (2021), etc). This literature analyzes the impact of inaccurately tracking external data on decision-making. The second subsystem, which I term the memory system, handles internal information — encoding, storing, and retrieving it as needed. I draw from Silveira et al. (2020) in modeling memory as a recursive system, where information must be actively encoded to remain accessible.

While attention and memory are complex mechanisms, I propose a simplified model that highlights their shared role in generating mental representations. The attention system creates a representation of external information, while the memory system represents internal knowledge. I posit that both systems operate with limited mental resources, affecting the accuracy of their representations. This separation draws inspiration from the fact that attention and memory systems are often associated with distinct brain regions. Despite functioning in parallel, these systems interact: attention guides what external information can be encoded into memory, and memory can subsequently influence the focus the attention. This interaction is crucial in understanding the forecast biases as I discuss in Section 2.

Specification of the representational systems. I propose to model mental representation as a linear-Gaussian filter of original information, whether external or internal. This representation is a noisy summary where the original information undergoes a linear transformation followed by the addition of Gaussian noise (orthogonal to the original information). Importantly, I assume the filter's specification arises from optimal cognitive processes designed to minimize expected losses from inaccurate forecasts. This implies that attention and memory systems are jointly determined to minimize the loss function (1.3), given the constraint that forecasts integrate information from both systems with limited accuracy.

1.2.1 Attention System: Mental Representation of External Information

Many pieces of publicly available information partially reveal the underlying state z_t. Examples include historical realizations of past y_t or other economic variables relevant for predicting z_t. All such information that is at least somewhat informative about the value of z_t can be stored in a large vector N_t. I further suppose that the relationship between N_t and z_t is described as follows:[1]

$$N_t = R \cdot z_t + \nu_t \tag{1.5}$$

where R is a constant vector, and $\nu_t \sim \mathcal{N}(0,\, V)$ for some positive definite matrix V. Forecasters are assumed to be correctly aware of both the structure and contents of N_t.

I suppose that the attention system generates the mental representation of N_t in the following form.

$$n_{i,t} = K_t \cdot N_t + u_{i,t} \tag{1.6}$$

Here, K_t is a matrix (possibly with many fewer rows than the number of elements in N_t) and $u_{i,t} \sim \mathcal{N}(O,\, \Sigma_{ut})$ for some positive semi-definite matrix Σ_{ut}. The noise $u_{i,t}$ is not correlated

1. Because of the Markov nature of the data-generating process, it is the information about z_t that is most relevant to DM. One could extend this set-up to incorporate a more complicated data-generating process.

with N_t and idiosyncratic to each forecaster. The specific forms of the matrices K_t and Σ_{ut} remain to be determined.

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 $\mathcal{I}\left(n_{i,t};\, N_t\right)$.[2] 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:[3]

$$I\left(n_{i,t}; N_t\right) \leq -\frac{1}{2}\log \phi_n \tag{1.7}$$

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 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 Σ_{ut} 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 Σ_{ut} are positive. When $\phi_n \to 1$, forecasts are not based on information in N_t, since the representation is infinitely inaccurate.

1.2.2 Memory System: Mental Representation of Internal Information

Through forecasting experiences, DM accumulates knowledge about the state of the economy. I suppose that DM's stock of knowledge at $t-1$ can be described with a vector $M_{i,t}$.

$$M_{i,t} = \begin{pmatrix} m_{i,t-1} \\ \\ n_{i,t-1} \end{pmatrix} \tag{1.8}$$

where $m_{i,t-1}$ denotes the knowledge carried through $t-1$ (before observing N_{t-1}) and $n_{i,t-1}$ is the knowledge from observing the news vector N_{t-1}. Thus, $M_{i,t}$ is the internal information that DM can access at time t.

I suppose that the memory system represents the internal information in the following

2. This metric captures how "close" $n_{i,t}$ is to N_t. If $\mathcal{I}\left(n_{i,t};\, N_t\right)$ 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.

3. 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 $\mathcal{I}(n_{i,t}; N_t | n_{i,t-1}, \cdots, n_{i,0})$. 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.

way, analogous to the way attention system represents the external information in (1.6).

$$m_{i,t} = \Lambda_t \cdot M_{i,t} + \omega_{i,t} \tag{1.9}$$

where Λ_t is a matrix that may have fewer rows than $M_{i,t}$ and $\omega_{i,t}$ is an i.i.d. sequence that is uncorrelated with $M_{i,t}$ and drawn from the Gaussian distribution $\mathcal{N}(O, \Sigma_{\omega,t})$ for some positive semi-definite matrix $\Sigma_{\omega,t}$. Again, the entire sequence of the two matrices Λ_t and $\Sigma_{\omega,t}$ is to be specified.

The extent of noise in the mental representation $m_{i,t}$ is measured with the Shannon mutual information between $m_{i,t}$ and $M_{i,t}$. The lower mutual information captures a more inaccurate representation of internal information. In parallel with (1.7), I assume that the accuracy of the representation is constrained as follows:

$$I(m_{i,t}; M_{i,t}) \leq -\frac{1}{2}\log\phi_m \tag{1.10}$$

for $\phi_m \in (0,1)$ taken as given. A higher ϕ_m means a more constrained representation. If ϕ_m approaches 0, the internal information encompasses all past observations $(n_{i,0}, n_{i,1}, \cdots, n_{i,t-1})$, in which case memory is perfectly nested (i.e., $M_{i,t} \supseteq M_{i,t-1}$). In contrast, when ϕ_m approaches 1, the mental representation of internal information solely reflects noise and offers no predictive values for the hidden state z_t.

1.3 Implications of the Linear-Gaussian Representational Systems

We have seen how external and internal information is mentally represented. For brevity, I refer to $n_{i,t}$ as *noisy news* (i.e., an imperfect representation of external information) and $m_{i,t}$ as *noisy memory* (i.e., an imperfect representation of internal information). I suppose that forecasters are skilled at integrating their cognitive states. As they optimally utilize $(m_{i,t}, n_{i,t})$, their forecasts align with principles of Bayesian efficiency. The conditional distribution is derived using the usual Kalman filter formula.

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.[4] 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 for

4. Note that the subjective probability distribution may not align with the true underlying probabilities. In particular, the second moment of the Gaussian distribution captures the uncertainty one perceives, which depends on the severity of attention and memory noise.

any τ implied by her cognitive states:

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

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} must be $E\left[\rho^h \left(z_t - z_{i,t|t}\right)^2\right]$, whose 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 thus proportional to $\Sigma_{z,t|t}$. The loss function (1.3) then reduces to

$$\sum_{t=0}^{\infty} \beta^t \left[q \cdot \Sigma_{z,t|t}\right], \tag{1.11}$$

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

1.4 The Optimal Cognitive Process

The sequence $\{K_t, \Sigma_{ut}, \Lambda_t, \Sigma_{\omega t}\}_{t=0}^{\infty}$ fully describe the cognitive process of accessing and storing information over time. I propose that the representational systems are designed to minimize the loss function defined in (1.11) subject to the information environment specified by equations (1.6), (1.7), (1.9), and (1.10). Thus, attention and memory systems are jointly determined. This formalization captures the essence of the optimal cognitive process, and its key insights remain valid even for more complex state spaces introduced in Section 3.

1.4.1 Optimal Representation of Noisy News

Though noisy news can take many different forms as seen in the equation (1.6), the optimal representation has a lower dimension compared to the raw external information (N_t in equation (1.5)). Additionally, the proposition below identifies the specific information from N_t that gets encoded in this mental representation.

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

$$\tilde{n}_{i,t} = \kappa_{zt} \cdot E\left[z_t | N_t\right] + \tilde{u}_{i,t} \tag{1.12}$$

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

Proof. See Appendix B. □

Why is the optimal $n_{i,t}$ is one-dimensional and has the structure in (1.12)? 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 $z_t | N_t$ follows a Gaussian distribution, such information is summarized in the first moment.[5] Therefore, $\tilde{n}_{i,t}$ encodes $E\left[z_t | N_t\right]$, which can be denoted as follows without loss of generality:

$$E\left[z_t | N_t\right] = z_t + \tilde{\nu}_t, \quad \tilde{\nu}_t \sim \mathcal{N}\left(0, \sigma_\nu^2\right) \tag{1.13}$$

Here, $\tilde{\nu}_t$ is the common errors in N_t, and its variance σ_ν^2 is taken as given and known to forecasters.

As one can see from (1.12), there are combinations of κ_{zt} and σ_{ut}^2 that imply the same posterior distribution $z_t | m_{i,t}, \tilde{n}_{i,t}$ for any given $m_{i,t}$. Therefore, I impose a normalization so that κ_{zt} alone captures the accuracy of the representation. In particular, I impose that $Cov\left[z_t, \tilde{n}_{i,t} | m_{i,t}\right] = Var\left[\tilde{n}_{i,t} | m_{i,t}\right]$, in which case the posterior uncertainty is determined as

$$\Sigma_{z,t|t} = \left(1 - \kappa_{zt}\right) \Sigma_{z,t|t}^m$$

for a given prior uncertainty $\Sigma_{z,t|t}^m$. That is, observing N_t and basing one's forecasts on $\tilde{n}_{i,t}$ reduces the uncertainty about z_t by a factor of $1 - \kappa_{zt}$. The normalization pins down σ_{ut}^2 as the following function of κ_{zt}:

$$\sigma_{ut}^2 = \kappa_{zt}\left(1 - \kappa_{zt}\right) \Sigma_{z,t|t}^m - \kappa_{zt}^2 \sigma_\nu^2 \tag{1.14}$$

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

Furthermore, the accuracy constraint as described in the equation (1.7) pins down the optimal value of κ_{zt}. When considering the optimal structure of $n_{i,t}$ from equation (1.12), the mutual information between $n_{i,t}$ and N_t becomes equivalent to $\mathcal{I}\left(\tilde{n}_{i,t};\, z_t + \tilde{\nu}_t\right)$. The value of this mutual information is entirely determined by κ_{zt}. Since the accuracy constraint imposes a limit on the highest achievable mutual information, it becomes clear that the optimal value

5. Note that the second moment, $Var\left[z_t | N_t\right]$, is a fixed value determined by the underlying data-generating process of z_t and the structure of the news vector. Consequently, a decision-maker faces no uncertainty regarding this moment.

for κ_{zt} is the highest possible one. Consequently, we can directly express κ_{zt} as the following function of ϕ_n.

$$\kappa_{zt} = \frac{\Sigma^m_{z,t|t}}{\Sigma^m_{z,t|t} + \frac{\phi_n}{1-\phi_n}\left(Var\left[z_t\right] + \sigma^2_\nu\right) + \sigma^2_\nu} \tag{1.15}$$

Noisier news corresponds to lower values of κ_{zt} , leading to increased uncertainty in the posterior estimate. After sufficient learning, both the prior and posterior uncertainties, represented by $\Sigma^m_{z,t|t}$ and $\Sigma_{z,t|t}$, reach a stable positive level, independent of time. This convergence results in a constant value for κ_{zt} as well, symbolized by $\kappa_{zt} \to \kappa_z$.

1.4.2 Optimal Representation of Noisy Memory

Just like we saw with noisy news, the optimal way to represent noisy memory is not necessarily by storing everything. While equation (1.9) suggests it could have any dimension, the optimal representation is more compact compared to the raw internal information, $M_{i,t}$ in equation (1.8)). As the following proposition explains, specific information from $M_{i,t}$ gets encoded in the mental representation.

Proposition 2. *$\tilde{m}_{i,t}$ is the optimal representation of $M_{i,t}$ such that*

$$\tilde{m}_{i,t} = \lambda_t \cdot E\left[z_t | M_{i,t}\right] + \tilde{\omega}_{i,t} \tag{1.16}$$

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. □

Building on the analysis of noisy news representation, we can derive the optimal structure for representing memories using a similar approach. Just like how the optimal noisy news representation, $\tilde{n}_{i,t}$, captured information from N_t useful for predicting the realized value of z_t, the optimal representation of noisy memory captures information from $M_{i,t}$ that is useful for predicting z_t. The critical information again boils down to the expected value of $z_t | M_{i,t}$. Therefore, $\tilde{m}_{i,t}$ essentially stores the average outcome of z_t one would expect based on the previous understanding (as represented by $z_{i,t|t-1}$). This connection highlights a parallel between how one processes information from the external world (news) and the internal world (memories): both prioritize capturing the gist, not every detail, to make accurate predictions about the state of the economy.

As one can see from (1.16), 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:

$$\sigma^2_{\omega t} = \lambda_t\,(1-\lambda_t)\, Var\left[z_{i,t|t-1}\right] \tag{1.17}$$

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, $\Sigma_{z,t|t-1}$, the prior uncertainty is determined as follows:

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

Uncertainty about z_t increases from $\Sigma_{z,t|t-1}$ to $\Sigma^m_{z,t|t}$ because prior knowledge is imperfectly represented when making new forecasts. In the extreme case of no-memory ($\lambda_t \to 1$). the prior uncertainty $\Sigma^m_{z,t|t}$ converges to the "default" (or initial) uncertainty $Var\left[z_t\right]$.

The optimal value of λ_t can be determined within the context of the memory system's accuracy constraint, formally defined in equation (1.10). This constraint effectively mandates a limitation on the fidelity of information retrievable from memory, which translates to maximizing the mutual information between the mental representation ($\tilde{m}_{i,t}$) and the summarized internal information ($z_{i,t|t-1}$). Consequently, the optimal value of λ_t becomes the one that maximizes this mutual information while adhering to the aforementioned constraint. The optimal value can be directly expressed as a function of ϕ_m, which encapsulates the inherent limitations of the memory system.

$$\lambda_t = 1 - \phi_m \tag{1.18}$$

A direct observation can be made that noisier memory corresponds to lower values of λ_t and concomitantly, higher prior uncertainty.[6] Since ϕ_m is a constant parameter, I use λ to denote the level of λ_t in the following section.

1.5 Discussion: Interpretations of information frictions

Traditional models of information frictions focus on how difficult it is to get information. For example, the "noisy" information model assumes forecasters only have fragmented information about the state of the economy (Woodford (2003)), while the "sticky" information model suggests some forecasters rely on outdated data (Mankiw and Reis (2002)). However, these explanations are less convincing for professional forecasters who typically have timely access to readily available data.

6. The determination of memory system is more complicated in the general case introduced in Section 3, but the intuition remains valid.

This paper proposes a different perspective: information processing constraints. That is, limitations exist in understanding the implications of the available information, not necessarily obtaining it. Forecasters are aware of current events, like labor strikes or the recent interest rate decisions by the central bank, but struggle to predict their exact impact on future inflation. This processing constraint leads to differing interpretations of the same data, even among professionals with equal access to information sources. The proposed model expands on Sims (2003) by applying the processing constraint to *all* information, not just observable data. In my proposal, relying on internal knowledge incurs a mental cost, similar to the cost of using external information. This results in an attention and memory system that is inherently inaccurate. In comparison, Sims (2003) assumes that the knowledge from observable data can be re-accessed with accuracy in the future.

2 Model Predictions about Forecast Biases

This section investigates the influence of cognitive noise on the forecasts of the hidden state z_t. The analysis reveals that cognitive noise systematically biases these forecasts, leading to patterns consistent with biases observed in professional forecasts of diverse macroeconomic and financial variables. I also propose an estimation strategy that leverages survey forecasts to quantify the extent of cognitive constraints.

2.1 Forecasts Subject to Cognitive Noise

According to the proposed framework, internal information is represented as described by equation (1.16). This equation informs how the DM's time-t prior belief about z_t will deviate from the perfect-memory scenario ($z^m_{i,t|t} = z_{i,t|t-1}$). Specifically, the prior beliefs takes the form:

$$z^m_{i,t|t} = (1-\lambda)\,E\left[z_t\right] + \,\lambda\, z_{i,t|t-1} \,+ \tilde{\omega}_{i,t} \tag{2.1}$$

Here, λ captures the inaccuracy of the memory system, while $\tilde{\omega}_{i,t}$ is the variability associated with memory noise, as defined in equations (1.18) and (1.17). The key takeaway is that when memory is imprecise ($\lambda < 1$), forecasts show sluggishness in incorporating one's past knowledge. This arises because the processing of internal information introduces noise, causing the remembered knowledge about z_t to be biased towards a default prior ($E\left[z_t\right]$).

Similarly, imprecise attention affects the evolution of the posterior belief. When external information arrives and mentally represented according to equation (1.12), the DM's prior belief updates to the following posterior belief:

$$z_{i,t|t} \;=\; (1-\kappa_z)\, z^m_{i,t|t} + \kappa_z\, z_t + \kappa_z\, \tilde{\nu}_t + \tilde{u}_{i,t} \tag{2.2}$$

where κ_z controls the weight given to new information, $\tilde{\nu}_t$ captures the "common noise" in the external information, and $\tilde{u}_{i,t}$ is the individual idiosyncratic attention noise. This equation reveals a different type of sluggishness compared to imperfect memory discussed earlier. Due to processing constraints, forecasts tend to put less weight on new information ($\kappa_z \ll 1$) and more on one's prior beliefs. This results in sluggish updates, meaning forecasts are slow to catch up with recent developments in z_t.

Equations (2.1) and (2.2) jointly describe the evolution of the DM's beliefs about z_t. These equations capture how cognitive noise manifests in the belief formation process.

$$z_{i,t|t} = (1-\lambda)\,(1-\kappa)\,E\left[z_t\right] + \lambda\,(1-\kappa_z)\, z_{i,t|t-1} + \kappa_z\, z_t + \kappa_z\, \tilde{\nu}_t + (1-\kappa_z)\,\tilde{\omega}_{i,t} + \kappa_z\, \tilde{u}_{i,t} \tag{2.3}$$

This equation summarizes the key features of forecasts under the influence of cognitive noise. Notably, limitations in processing external information, captured by the value of κ_z lower than the perfect-attention scenario, impede the timely incorporation of changes in z_t, leading to sluggish recognition of new developments and delayed adaptation of forecasts. Furthermore, memory constraints, represented by the value of λ lower than unity, hinder DM's ability to swiftly integrate past knowledge, resulting in sluggish updates of forecasts based on historical experience. Finally, individual variations in processing information (reflected in $\tilde{\omega}_{i,t}$ and $\tilde{u}_{i,t}$) generate forecast dispersion. This implies that even with access to the same sources of external information, individuals arrive at different predictions.

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

$$z_{i,t|t} = (1-\kappa^*_z)\, z_{i,t|t-1} + \kappa^*_z\, z_t + \kappa^*_z\, \tilde{\nu}_t + +\kappa^*_z\, \tilde{u}_{i,t} \tag{2.4}$$

where κ^*_z is the limit of the Kalman gains such models. Comparing this law of motion to the equation (2.3) in the noisy memory case, three key changes stand out. With noisy memory, less weight is given to prior knowledge due to the term $\lambda < 1$. This implies that past information has a reduced influence on current beliefs. Second, the weight accorded to new information, represented by κ_z, is larger than κ^*_z in equation (2.4). This indicates that with noisy memory, individuals place more emphasis on newly received information. Finally, a new source of cognitive noise emerges in the form of $\tilde{\omega}_{i,t}$ in equation (2.3). This additional noise further disrupts the accuracy of forecasts.

Figure 1 illustrates the effects of noisy memory when learning about z_t. I use the parameter values $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$ for the data-generating process, and $\sigma_\nu^2 = 0.2 \times \sigma_z^2$ for this numerical exercise. The top panel depicts how the extent of noisy news (ϕ_n) and noisy memory (ϕ_m) affect the level of Kalman gain. I categorize the pairs of noise parameters as either "overreaction" or "underreaction", depending on whether the resulting Kalman gain is greater or smaller than the benchmark case of ($\phi_n = 0, \phi_m = 0$). As conventional models of information frictions suggest, the Kalman gain is lower than the benchmark when the only cognitive noise is the noisy news. On the other hand, the opposite is true when only noisy memory is present. When both noises are present, a spectrum of under- to overreaction exists. The relative strength of overreaction (driven by noisy memory) and underreaction (driven by noisy news) determines the size of the Kalman gain. For a given level of noisy memory, we observe underreaction for a sufficiently small level of ϕ_m, but a larger ϕ_m can flip it overreaction.[7]

The bottom panel shows the impulse response to innovation in z_t. The grey dotted line shows the response of z_t itself. Other lines show the response of forecasts of z_t for varying degrees of cognitive noise. The black solid line is the benchmark case without any cognitive noise. The forecasts lag the actual z_t because even an efficient use of external information does not accurately reveal the hidden state. The blue dashed line describes the forecasts when only noisy news is present. This line initially undershoots the benchmark but catches up with enough learning opportunities. In comparison, the orange dashed-dotted line shows the case when both noises are present, in particular when no prior knowledge is recalled at all. The parameter values chosen for this line correspond to the "overreaction" region in the above figure. Due to the high prior uncertainty resulting from no memory, the initial response is more significant than the benchmark. Such high level of uncertainty arises because accumulation of knowledge is slow (in this case, zero). Since DM cannot tap into prior knowledge, DM's knowledge about z_t does not improve over time learning about z_t, even with the large Kalman gain.

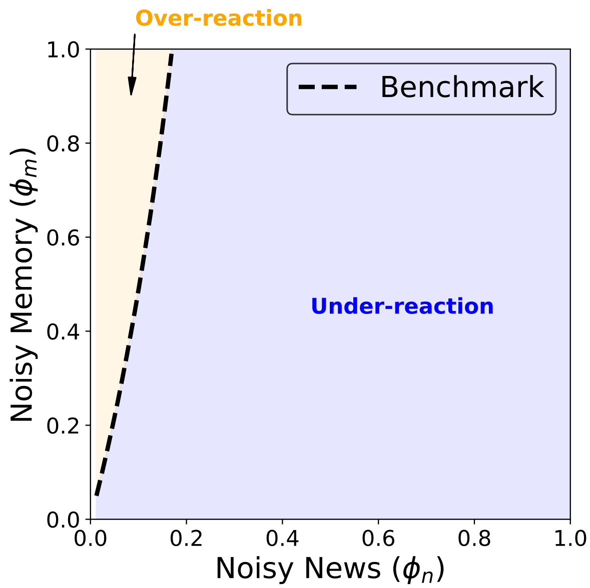
2.2 Biases in Survey Forecasts

This section dives into two regression tests designed to probe the potential deviations of survey forecasts from the predictions of the Full-Information Rational Expectations (FIRE) hypothesis. By analyzing these deviations, I aim to shed light on the possible presence and

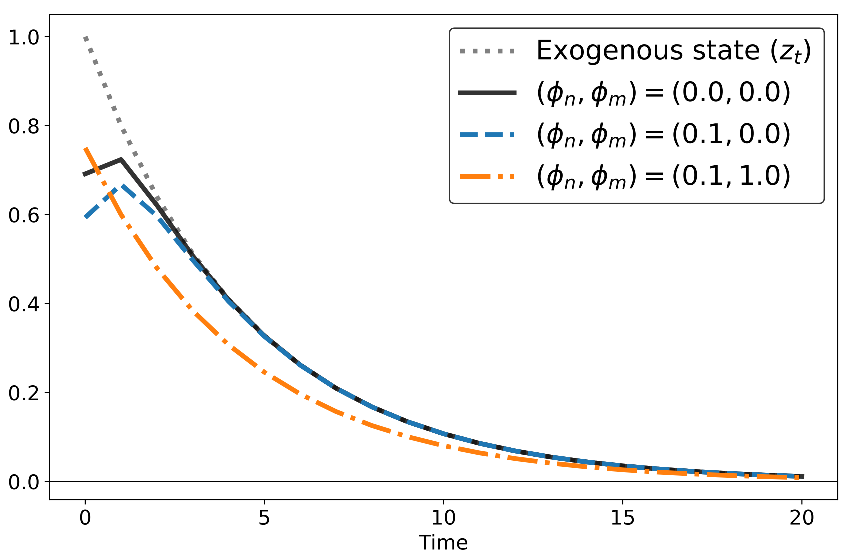
7. One factor that crucially determines the exact size of each region is σ_ν^2, the prevalence of the noise in external information, as it determines the no-noise benchmark Kalman gain. If the external information completely reveals the hidden state, then the benchmark Kalman gain is one; in this extreme case, any predicted size of gains subject to cognitive noise points to underreaction.

Figure 1: The effects of noisy memory

(a) Kalman gains



(b) Impulse responses



The figures describe the effects of noisy memory. The top panel categorizes pairs of noisy news (ϕ_n) and noisy memory (ϕ_m) as "overreaction" or "underreaction" regions, depending on whether the resulting Kalman gains are greater or smaller than the benchmark free of cognitive noise. The bottom panel depicts the impulse response of forecasts to an innovation in z_t. The grey dotted line is the response of z_t. The black solid line shows the case of perfect news and memory. Colored lines assume a varying degree of cognitive noise. The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$, and the extent of noise in external information is quantified as $\sigma_\nu^2 = 0.2 \times Var\left[z_t\right]$.

influence of cognitive noise in forecasting behavior.

Before delving into the analysis of survey forecasts, it is useful to establish the three fundamental assumptions that underpin the FIRE hypothesis. We can view these assumptions through the lens of the model established in Section 1. First, the Bayesian efficiency assumption states that agents optimally utilize all available information.[8] Within the proposed model, this implies that agents update their beliefs in a statistically optimal manner, incorporating all relevant information. In simpler terms, agents constantly minimize errors in their predictions. Second, the perfect memory assumption demands that agents have flawless recollection of all past information. In the context of the proposed model, it signifies that agents can effortlessly recall and leverage past knowledge with complete accuracy when forming new beliefs. This allows them to draw upon their entire history of learning and experiences. Finally, the perfect attention assumption posits that agents can process all available external sources of information needed for optimal belief updating. Within the proposed model, this translates to every agent having access to the same perfect external information when making predictions.

The three core assumptions of FIRE - Bayesian efficiency, perfect memory, and perfect attention - pave the way for a series of increasingly restrictive regression tests. Each test unveils another layer of potential deviations from FIRE. To start with, the Bayesian efficiency assumption implies that individual forecast errors cannot be systematically predicted by any information readily available to the forecaster. This assumption can be tested by regressing forecast errors on the element of the information set.

Building upon this first test, perfect memory assumption introduces the additional constraint that past forecast revisions should not be predictive of future errors. Since these revisions are part of the forecaster's information set, they should not yield any insights into future mistakes if perfect memory holds true. Testing this assumption can be done by estimating the following regression.

$$y_{t+h} - F_{i,t}\, y_{t+h} = \alpha_I + \beta_I \left(F_{i,t}\, y_{t+h} - F_{i,t-1}\, y_{t+h}\right) + e_{i,t+h|t}, \tag{2.5}$$

where $F_{i,t}\, y_{t+h}$ and $F_{i,t-1}\, y_{t+h}$ is forecaster i's predictions of y_{t+h} at time t and $t-1$, respectively. The test assesses the combined influence of Bayesian efficiency and perfect memory.

Finally, perfect attention implies that errors in average forecasts (across multiple individuals) would not be predictable by recent revisions in the average forecasts. This is because, under perfect attention, all relevant information is incorporated into all individuals' forecasts, and thus into the average forecasts. This idea is captured by the regression test of the

8. Agents are assumed to possess accurate knowledge of the economic environment, as was the case in the model in section 1.

following form.

$$y_{t+h} - F_t\, y_{t+h} = \alpha_C + \beta_C\,(F_t\, y_{t+h} - F_{t-1}\, y_{t+h}) + e_{t+h|t}, \tag{2.6}$$

where $F_t\, y_{t+h}$ and $F_{t-1}\, y_{t+h}$ are the average predictions at time t and $t-1$ among all forecasters. This test probes the combined influence of all three assumptions at the aggregate level. Scrutinizing survey forecasts through these progressively stricter tests help unveil layers of possible deviations from FIRE.

Crucially, these two regression tests hold particular significance in the expectation formation literature. They offer a structured methodology to assess whether the concept of information frictions can explain the observed biases in survey forecasts. Coibion and Gorodnichenko (2015) argue that a violation of test (2.6) aligns with the popular models of information frictions. Conversely, Bordalo et al. (2020) posit that a violation of test (2.5) cannot be explained by information frictions; instead, the authors argue that it further requires a relaxation of Bayesian efficiency. I discuss these arguments in detail and offer different interpretations of the regression coefficients based on the model proposed in this paper.

Interpretations of β_C. Coibion and Gorodnichenko (2015) analyze the relationship between forecast errors and revisions of the average forecasts, finding a positive and statistically significant coefficient (β_C) for various macroeconomic variables. They attribute this finding to the relaxation of the full-information assumption, arguing that information frictions can lead to sluggish forecast revisions. Intuitively, when individuals lack complete information, their forecasts may incorporate unrelated noise or outdated data. This can create inertia in aggregate forecast revisions, when new economic changes emerge. Consequently, forecast errors, defined as in the dependent variable in test (2.6), become positively correlated with revisions, as sluggish adjustments lead to persistent deviations from the true realizations.

The authors contend that the magnitude of β_C reflects the severity of information frictions. In their framework, a larger β_C could stem from either a higher noise level in the information signal (from the noisy information model) or a longer information lag (from the sticky information model). This implies that agents with more severe frictions experience greater difficulty in extracting or incorporating new information in their forecasts, leading to a stronger positive association between forecast errors and revisions.

The cognitive-noise expectation model introduced in section 1 offers a novel interpretation of the regression coefficient β_C. Notably, the analysis departs from prevalent models of information frictions that often implicitly assume perfect recall of past knowledge. The following proposition provides a detailed examination of how imperfect memory shapes the regression coefficient and influences its interpretation.

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

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

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 $\phi_n \leq \bar{\phi}_n \equiv \bar{g}(\rho, \sigma_\epsilon^2)$.*

Proof. See Appendix C. □

Noisy news and noisy memory jointly influence the Kalman gain (κ_z), impacting the dynamics of forecast revision captured by β_C. The first part of the proposition confirms that findings in Coibion and Gorodnichenko (2015) still hold true when the perfect-memory assumption is relaxed: Due to imperfect awareness of the true state caused by noisy news, forecasters update their beliefs sluggishly, leading to a positive β_C. Furthermore, noisier news reduces κ_z and increases β_C.

However, this model offers a novel insight: the combined effect of noisy news and noisy memory on β_C. Noisy memory implies less accurate prior knowledge, leading to higher uncertainty about the state. Consequently, forecasters rely more heavily on incoming data, reflected in a higher κ_z. This results in a *lower* β_C.

Interpretations of β_I. Bordalo et al. (2020) empirically challenge the findings of Coibion and Gorodnichenko (2015) by observing a a negative β_I across numerous macroeconomic and financial variables. This result contradicts theoretical predictions of a purely information-based frictions; regardless of the extent of information frictions, β_I should be zero, as long as forecast revisions are incorporated within the forecaster's information set. Thus, Bordalo et al. (2020) contend that information frictions cannot solely explain the observed forecast bias and suggest the need to relax the assumption of Bayesian efficiency. To accommodate this finding, the authors propose a non-Bayesian model termed "diagnostic expectations." Within this framework, forecasters overweight recent observations, leading to excessive forecast revisions. The authors posit that a more negative β_I signals a great departure from Bayesian efficiency.

In contrast, I have proposed a framework that relaxes the assumption of perfect memory while maintaining Bayesian efficiency. This model offers an alternative interpretation of the regression coefficient, as outlined in the following proposition.

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

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

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 C. □

This proposition demonstrates that imperfect memory, even within the bounds of Bayesian efficiency, can generate a negative β_I. The regression coefficient, within my framework, reflects an inherent bias due to the under-utilization of one's past knowledge.

Importantly, the degree which forecasters leverage past knowledge is directly influenced by the interplay of noisy news and noisy memory. Noisy memory leads to the discounting of prior knowledge, manifesting in a negative β_I. This results from the imperfect recall of past knowledge. The effect is exacerbated when processing external news becomes more difficult. In this scenario, forecasters lean more heavily to their imprecise memory as a foundation for their judgments. With external information less able to correct memory-driven biases, β_I becomes increasingly negative.

Identifying the extent of cognitive constraints. Propositions 1 and 2 demonstrate that the two regression coefficients allow us to quantify the severity of noisy news and noisy memory.

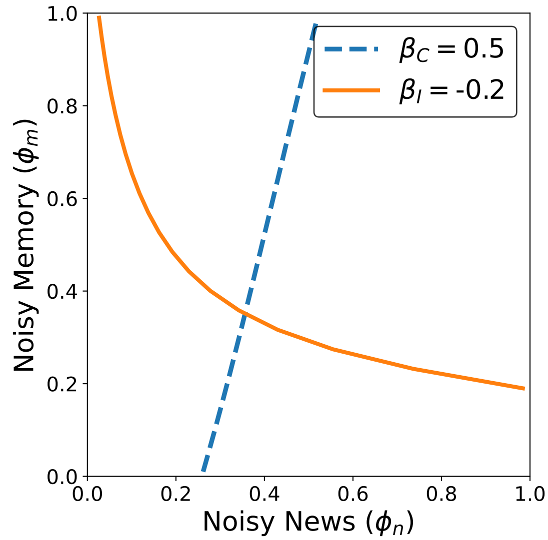
Lemma 1. *Given levels of β_C and β_I identify a unique pair of ϕ_n and ϕ_m, if it exists.*

Proof. See Appendix C. □

This lemma establishes a one-to-one mapping between the regression coefficients (β_C, β_I) and the underlying parameters (ϕ_n, ϕ_m). The key idea is to construct isocurves: sets of ϕ_n and ϕ_m that are consistent with a given coefficient value. This illustrates the influence of both noisy news and noisy memory in shaping the forecast biases captured by β_C and β_I. I demonstrate that the isocurve of β_C exhibits a positive slope, while the isocurve of β_I exhibits a negative slope. Thus, if they intersect, they do so at a single point.

Figure 2 illustrates this point using the parameters of $\rho = 0.8$, $\sigma_\epsilon^2 = 1.0$, and $\sigma_\nu^2 = 0.2\sigma_z^2$. The isocurve for $\beta_C = 0.5$ (solid blue) slopes upward, indicating that increased belief sensitivity due to noisier memory can be offset by slower updates due to noisier news.

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



This figure shows the isocurves for the two regression coefficients in (2.6) and (2.5). 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, which is $\phi_n^* = 0.35$ and $\phi_m^* = 0.35$. The data-generating process is described by $\rho = 0.8$ and $\sigma_\epsilon^2 = 1$, and the extent of noise in external information is quantified as $\sigma_\nu^2 = 0.2 \times \sigma_z^2$.

The isocurve for $\beta_I = -0.2$ (dashed orange) slopes downward, showing that greater underutilization of past knowledge (due to noisier memory) can be offset by less reliance on memory when news is easier to process. Their interaction uniquely identifies the extent of noisy news and noisy memory that align with the observed regression coefficients as $\phi_n^* = 0.35$ and $\phi_m^* = 0.35$.

3 Model Implications: Term-Structure of Expectations

The proposed model of noisy mental representation has a new implication about the term-structure of expectations. I extend the model introduced in 1 to reflect that forecasters face uncertainty about the long run. The predictions are consistent with several pieces of empirical evidence.

3.1 Learning about the Long Run

I consider the possibility that forecasters face uncertainty about the very long run. In this section, I suppose that the data-generating process of z_t is described as

$$z_t - \mu_t = \rho\,(z_{t-1} - \mu_{t-1}) + \epsilon_t \tag{3.1}$$

$$\mu_t = (1 - \rho_\mu)\,\mu + \rho_\mu\,\mu_{t-1} + \epsilon_{\mu,t}. \tag{3.2}$$

Compared to section 1, the exogenous state z_t fluctuates around the mean that also has stochastic components. Given the nature of external information, as specified in equation (1.5), DM thus has imperfect awareness of the actual realizations of both z_t and μ_t. In other words, the state variable relevant for predicting future realizations is expanded from z_t to (μ_t, z_t), as forecasts for y_{t+h} depend on DM's beliefs about μ and z_t. I denote this state vector as

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

I suppose that the information environments are assumed to be the same as in Section 1. That is, forecasters have access to a commonly available news source (1.5), which is mentally represented in their mind as described in (1.6). Each forecaster also taps into one's internal information (1.8), which has mental representation of the form (1.9). Any such representational system implies that forecasters' beliefs about the state vector x_τ for any τ are described as Gaussian distributions. I use the following notations to describe them.

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

As was the case in Section 1, the first moment will depend on the history of realized cognitive noise and thus is denoted with a subscript i. The second moment evolves as a deterministic function of time, i.e. the length of the learning experiences.

The exact specification of the representational system is assumed to be the optimal one minimizing the objective function (1.3), which reduces to

$$\sum_{t=0}^{\infty} \beta^t\, trace\left(\Sigma_{t|t}\, Q\right), \tag{3.3}$$

where Q is a matrix defined as $Q \equiv \sum_{h=1}^{H} \alpha_h\, \alpha_h'$ and $\alpha_h = (1 - \rho^h \quad \rho^h)'$. As outlined in Section 1, not all representational systems described in equations (1.6) and (1.9) are feasible. They are subject to the constraints that the mental representation of both external and internal information has limited accuracy, as described in equations (1.7) and (1.10).

3.2 The Optimal Cognitive Process

I discuss the kind of information optimally encoded in the representational system. Unsurprisingly, the intuition carries over from Section 1 that some information does not improve the forecast accuracy while using up the mental resources, which therefore will not be part of the mental representation. As I derived in Appendix A, the optimal structure of noisy news and noisy memory takes the following analogous forms.

$$\tilde{n}_{i,t} = \tilde{K}_t \cdot E\left[x_t | N_t\right] + \tilde{u}_{i,t}, \quad \tilde{u}_{i,t} \sim \mathcal{N}\left(O, \Sigma_{ut}\right) \tag{3.4}$$

$$\tilde{m}_{i,t} = \tilde{\Lambda}_t \cdot E\left[x_t | M_{i,t}\right] + \tilde{\omega}_{i,t}, \quad \tilde{\omega}_{i,t} \sim \mathcal{N}\left(O, \Sigma_{\omega t}\right) \tag{3.5}$$

where both $\tilde{u}_{i,t}$ and $\tilde{\omega}_{i,t}$ are idiosyncratic random errors uncorrelated with $E\left[x_t | N_t\right]$ and $E\left[x_t | M_{i,t}\right]$, respectively. The sequence of matrices $\tilde{K}_t$, Σ_{ut}, $\tilde{\Lambda}_t$, and $\Sigma_{\omega t}$ for all t remain to be specified. The key message is that it is only the information about the state vector x_t stored in each information source that matters for minimizing the expected losses from not making accurate forecasting. This is a direct extension of the findings in the optimal representational systems, described in equations (1.12) and (1.16), when the only state variable is z_t.

Given this structure of the representational system, forecasters' beliefs are derived to evolve as follows. The beginning-of-period prior (based on noisy memory) evolves from the previous period's posterior distribution according to

$$x^m_{i,t|t} = x_{i,t|t-1} + \left(I - \tilde{\Lambda}_t\right)\left(E\left[x_t\right] - x_{i,t|t-1}\right) + \tilde{\omega}_{i,t} \tag{3.6}$$

$$\Sigma^m_{t|t} = \Sigma_{t|t-1} + \left(I - \tilde{\Lambda}_t\right)\left(Var\left[x_t\right] - \Sigma_{t|t-1}\right) \tag{3.7}$$

And the posterior belief (that is also based on noisy news) is described as

$$x_{i,t|t} = \left(I - \tilde{K}_t\right) x^m_{i,t|t} + \tilde{K}_t\, x_t + \tilde{\nu}_t + \tilde{u}_{i,t} \tag{3.8}$$

$$\Sigma_{t|t} = \left(I - \tilde{K}_t\right) \Sigma^m_{t|t} \tag{3.9}$$

where $\tilde{\nu}_t$ is common across forecasters and drawn from the distribution, $\tilde{\nu}_t \sim \mathcal{N}\left(O, \Sigma_\nu\right)$. Again, this term exists because the news vector does not completely reveal the realized value of the hidden state, as described in equation (1.5). Below I discuss how the optimal sequence of $\left\{\tilde{K}_t, \Sigma_{ut}, \tilde{\Lambda}_t, \Sigma_{\omega t}\right\}_{t=0}^{\infty}$ is derived.

Specifying $\tilde{K}_t$. As analyzed in section 3, the optimal representation is one-dimensional when forecasters are mainly uncertain about the current hidden state z_t (that is, they do not face additional uncertainty about the long run). In this case, the two scalars, κ_{zt} and σ_{ut}^2, fully describe the structure of the attention system, and their values are derived in equations (1.14) and (1.15). This system is closely related to the representational system optimal for the expanded state space. This is because the external information described in (1.5) is informative about the long-run μ_t only through its association with z_t. In other words, information in N_t about the additional state variable μ_t is subsumed in $E\left[z_t \middle| N_t\right]$. Specifically, the following expressions show how the matrices $\tilde{K}_t$ and Σ_{ut} in equation (3.4) are spanned from κ_{zt} and σ_{ut}^2.

$$\tilde{K}_t = \kappa_{zt} \cdot \frac{\Sigma_{t|t}^m \, e_2 \, e_2'}{e_1' \, \Sigma_{t|t}^m \, e_2} \tag{3.10}$$

$$\Sigma_{ut} = \sigma_{ut}^2 \cdot \frac{1}{\left(e_1' \, \Sigma_{t|t}^m \, e_2\right)^2} (\Sigma_{t|t}^m \, e_2)(\Sigma_{t|t}^m \, e_2)' \tag{3.11}$$

where the vectors e_1 and e_2 are defined as $e_1 \equiv (1 \quad 0)'$ and $e_2 = (0 \quad 1)'$.

Specifying $\tilde{\Lambda}_t$. Deriving the optimal memory system under an expanded state space introduces two key challenges. Firstly, the dimension of information to be retained over time increases. This necessitates a forward-looking optimization approach that explicitly considers the potential availability of external information sources in future periods. Secondly, and from a technical standpoint, the loss function exhibits a non-convexity with respect to the choice variables – the sequence of matrices defining the memory system across all periods. Consequently, deriving the optimal solution requires global optimization methods. To isolate the trade-off stemming from the first challenge, I consider the myopic case ($\beta \to 0$). This simplification enables the analytical derivation of the optimal $\tilde{\Lambda}_t$ in equation (3.5).

A crucial step in this derivation is defining a matrix Γ_t, which captures how the memory system differentially weights information components. This matrix is defined as follows:

$$\Gamma_t = \left(I - \tilde{K}_t\right)' Q \left(I - \tilde{K}_t\right) \tag{3.12}$$

Note that Γ_t is a function of $\tilde{\Lambda}_t$ since $\tilde{K}_t$ itself is a function of $\tilde{\Lambda}_t$ (through its influence on $\Sigma_{t|t}^m$). The structure of Γ_t reflects the relative importance assigned to different information components. This interpretation is driven by two key factors. First, large elements within the matrix $(I - \tilde{K}_t)$ indicate that external information sources offer limited resolution of uncertainty about the state. In such scenarios, the representational system places a higher premium on accurate memory to guide its forecasts. Second, the loss function matrix Q quantifies the cost of forecasting errors across different dimensions. The memory system prioritizes retaining and utilization information where such errors are most costly.

The derivation of optimal $\tilde{\Lambda}_t$ clarifies the principles by which the memory system prioritizes information and may strategically reduce its representational complexity. This complexity is directly reflected in the rank of the matrix $\tilde{\Lambda}_t$ as I discuss below. While I focus on the essential insights, see Appendix A for a full derivation. The optimal configuration of $\tilde{\Lambda}_t$ depends crucially on the matrix Γ_t, defined in equation (3.12). To analyze its structure, I employ eigen-decomposition:

$$Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\Gamma_t\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}} = U_t\,G_t\,U_t'$$

where U_t is an orthonormal matrix storing eigenvectors, and G_t is a diagonal matrix storing eigenvalues in descending order (that is, $g_{1,t} > g_{2,t}$). This process extracts and ranks the key informational components embedded within Γ_t. The eigenvector corresponding to the higher eigenvalue captures the variation in $Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\Gamma_t\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}$ that is more significant in minimizing the loss function (3.3). The optimal $\tilde{\Lambda}_t$ is then derived as follows:

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

where D_t is a diagonal matrix, whose elements emerge from the analysis of Γ_t. Specifically, it takes the following form:

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

This derivation clarifies the process by which the rank and the content of the matrix $\tilde{\Lambda}_t$ is determined. We observe two potential cases: the first is when the memory system stores only some dimensions of the information within $x_{i,t|t-1}$, thus reducing its overall rank. Here, the first diagonal element in D_t receives the maximal possible weight permitted by the memory constraint, while the second element is zero. This happens when the extent of memory constraint, ϕ_m, exceeds the ratio of eigenvalues, $g_{2,t}/g_{1,t}$. Intuitively, if much of the variation in $Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}\,\Gamma_t\,Var\left[x_{i,t|t-1}\right]^{\frac{1}{2}}$ is explained by the first component of the eigen-decomposition, then $g_{1,t}$ is exceedingly larger than $g_{2,t}$ (for a given ϕ_m). In this case, putting the maximal weight on the first component is optimal in minimizing the loss function. The second case arises when the rank of the remembered knowledge remains unchanged from that of $x_{i,t|t-1}$ (that is, the optimal memory state $\tilde{m}_{i,t}$ is two-dimensional). In this instance, the first diagonal element in D_t is larger than the second, implying that the variation corresponding to $g_{2,t}$ is assigned a smaller, but non-zero, weight by the memory system. In both cases, it is straightforward to verify that $\det\left(I-\tilde{\Lambda}_t\right) = \phi_m$, confirming that the memory constraint outlined in equation (1.10) is respected.

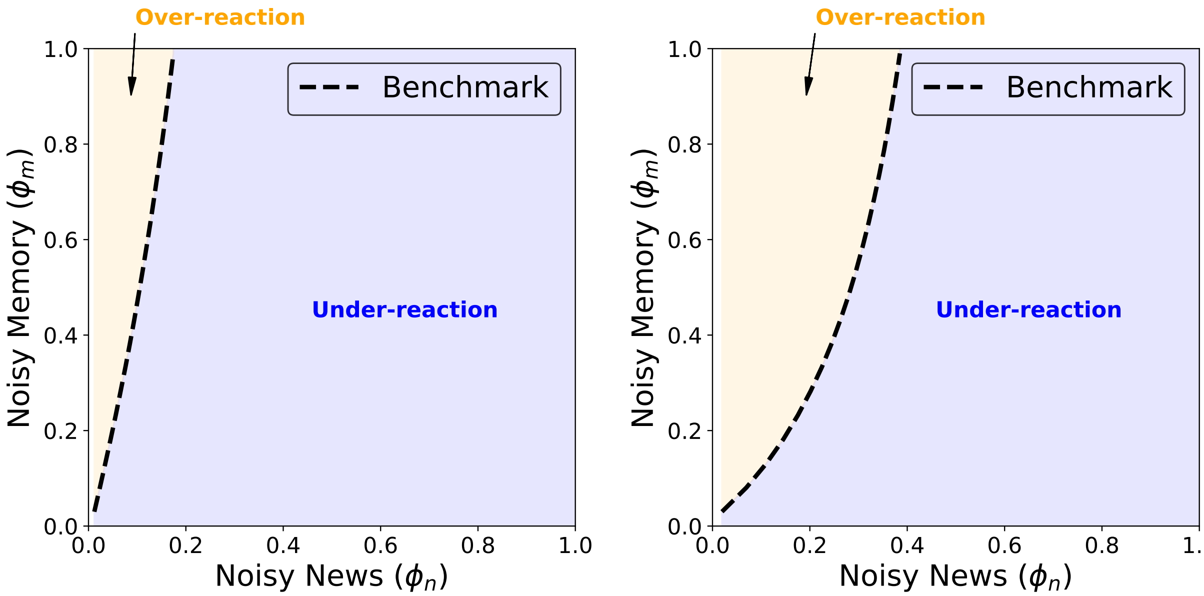
The principles in which $\tilde{\Lambda}_t$ is determined are similar to the core idea of the Principal Component Analysis in how factors are formulated and which ones are prioritized. The crucial difference in the exact determination of the factors lies in the role of memory system: it complements the external information available to DM in minimizing the future forecast errors.

3.3 Horizon-dependent Forecast Sensitivity

The joint consideration of both types of cognitive noises predicts that forecast sensitivity to news depends on the forecast horizon. I discuss this prediction in the context of the predicted values of Kalman gains when updating beliefs about the near-term and the long-term. Figure 3 again categorizes the pairs of cognitive noises as overreaction if the resulting Kalman gains are larger than the no-cognitive-noise benchmark, and underreaction otherwise. The parameters used to create the figures are: $\rho = 0.8$, $\sigma_\epsilon^2 = 1.0$, $\rho_\mu = 0.95$, and $\sigma_\mu^2 = 0.5$ for the data-generating process, and the extent of noise in external information is quantified as $\sigma_\nu^2 = 0.2 \times Var[z_t - \mu_t]$. The left panel illustrates the gains when updating the near-term forecasts. As discussed in section 2.1, underreaction is dominant for near-term. overreaction is possible, but only when the extent of noisy news is sufficiently modest. In comparison, the region of overreaction is larger for long-term forecasts. This has to do with the relative magnitudes of the benchmark Kalman gain values.

Given the trade-off between noisy news and noisy memory, the relative dominance of the two depends on the benchmark Kalman gains (without any cognitive noise). For near-term forecasts, the benchmark Kalman gains tend to be larger because external information is more directly relevant and informative about the near-term state of the variable being forecast. Since the benchmark near-term gains are larger to begin with, it requires larger degrees of noisy memory to push the Kalman gains above this higher benchmark level and into the overreaction region. In contrast, for long-term forecasts, the benchmark Kalman gain values are smaller because the external information is less directly informative about the more distant long-term state. With these smaller long-term benchmark gains, relatively smaller levels of noisy memory are sufficient to cause overreaction by pushing the gains above the benchmark. Therefore, it is likely that we observe underreaction from near-term forecasts and overreaction from long-term forecasts.

Figure 3: Kalman gains and forecast horizons



(a) When updating *near-term* forecasts (b) When updating *long-term* forecasts

The figures categorize pairs of noisy news (ϕ_n) and noisy memory (ϕ_m) as "overreaction" or "underreaction" regions, depending on whether the resulting Kalman gains are greater or smaller than the benchmark free of cognitive noise. The left panel illustrates the case for near-term forecasts, and the right panel is for long-term forecasts. The data-generating process is described by $\rho = 0.8$, $\sigma_\epsilon^2 = 1.0$, $\rho_\mu = 0.95$, and $\sigma_\mu^2 = 0.5$, and the extent of noise in external information is quantified as $\sigma_\nu^2 = 0.2 \times Var[z_t - \mu_t]$.

Inspecting the Mechanism

To further clarify the intuition, I consider the case where the uncertainty about the long-run originates not from the stochasticity, but from the parameter uncertainty. That is, the long-run trend of the exogenous state z_t is constant ($\mu_t = \mu$), and forecasters are uncertain about the exact level of μ. Suppose forecasters are correctly aware that the long-run is constant and learn about it, given the following Gaussian prior about μ

$$\mu \sim \mathcal{N}(\bar{\mu}, \Omega). \tag{3.13}$$

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

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

where c is a constant determined by the information environment. We can see that the precision of knowledge linearly increases in time, and the uncertainty eventually converges to zero after a long learning period.[9]

However, even when forecasters observe reasonably long data series, it is not guaranteed that they eventually face zero uncertainty about μ. As discussed in Silveira et al. (2020), forecasters fail to ever reach complete awareness of the model parameter when they have imperfect access to their prior knowledge due to memory frictions.[10] With noisy memory, DM imperfectly accesses internal information, and $Var\left[\mu \middle| m_{i,t}, n_{i,t}\right]$ does not converge to zero even after a long learning period. Intuitively, it is harder to accumulate knowledge when forgetful.

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 z_t is high, the DM partly attributes it to higher-than-expected μ and expects future z_t to be persistently high.[11]

Impulse response function. Figure 4 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 $\phi_n = 0.2$, $\phi_m = 0.2$, and $\Omega = 0.5 \times Var\left[z_t\right]$. 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 shows how forecasts for z_t evolve in response to the innovation. As was the case when only learning about z_t, learning about z_t is still sluggish because of noisy news. 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 4 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} so that it is easier to see whether forecasts undershoot or overshoot compared to the black dashed line. We see initial undershooting for all values of Ω because of 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

9. This relatively fast speed of learning motivates the assumption that economic agents are perfectly aware of the parameters describing the model environment. However, this assumption is not innocuous when learning takes a long time. This will be the case when there are simply not enough observations to be made to have the clarity about the parameters, say because disasters happen only so often (Collin-Dufresne et al. (2016)) or because the long-run trend has a complicated data-generating process unknown to forecasters (Farmer et al. (2024)).

10. Similar intuition follows from the work of Nagel and Xu (2022).

11. This prediction is similar to the extrapolative expectation models in the finance literature. Silveira et al. (2020) argues that limited memory might be the reason such extrapolation occurs.

Angeletos et al. (2021).[12]

Error-revision regression. The perpetual uncertainty about the long run also implies that the regression coefficients in the forecast error-revision test (2.6) and (2.5) will not be constant for different forecast horizons. To see why, consider the regression coefficient applied to forecasts for μ. Denoting the mean forecasts as $\hat{\mu}_{i,t} \equiv E\left[\mu | m_{i,t}, n_{i,t}\right]$ and the average forecasts as $\hat{\mu}_t \equiv \int \hat{\mu}_{i,t} d\, i$, we can see that

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

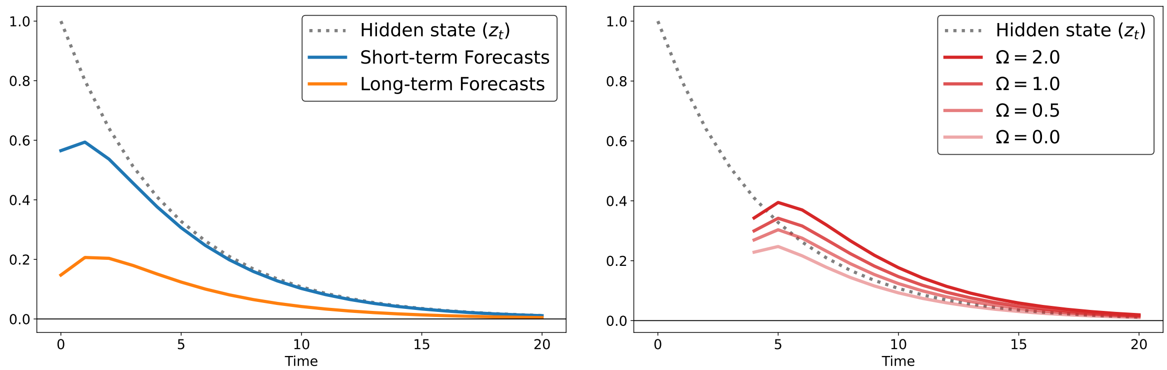
The derivation is straightforward: one can observe that β_C^{μ} must equal $-\frac{Var[\hat{\mu}_t|\mu] - Cov[\hat{\mu}_t, \hat{\mu}_{t-1}|\mu]}{2(Var[\hat{\mu}_t|\mu] - Cov[\hat{\mu}_t, \hat{\mu}_{t-1}|\mu])} = -\frac{1}{2}$. The same reasoning applies to deriving the value of β_I.[13] Forecasters revise their views about μ although μ is a fixed parameter.

Figure 5 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. All other model parameters follow the top panel of Figure 4. 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 $-\frac{1}{2}$. This pattern is in line with d'Arienzo (2020) and Wang (2021) that analyze professional forecasters' projections of interest rates. Both authors find that longer-horizon forecasts feature more negative biases when the regressions (2.6) and (2.5) are estimated. Bordalo et al. (2019) and Bordalo et al. (2023) find a similar pattern for stock analysts' forecasts for companies' long-term earnings.

12. 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).

13. Derivations for other horizons are in Appendix E.

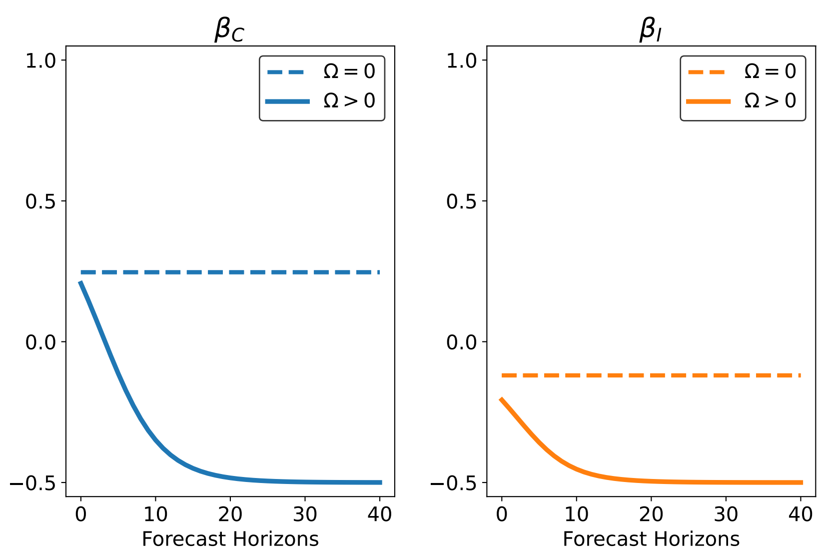
Figure 4: Impulse-response functions when learning about the long run



(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 extent of cognitive noise is set as $\phi_n = 0.2$ and $\phi_m = 0.2$. The top panel shows the response of z_t and the forecast of z_t and μ, when the initial uncertainty about μ is set as $\Omega = 0.5 \times Var[z_t]$. The bottom panel shows the response of four-period-ahead forecasts ($z_{i,t+4|t}$). Different lines assume varying degrees of Ω. Remaining parameters for the model environments are the same as in Figure 1.

Figure 5: β_C and β_I when learning about the long run



This figure shows model predictions of the two regression coefficients in (2.6) and (2.5) for different forecast horizons. The extent of cognitive noise is set as $\phi_n^* = 0.35$ and $\phi_m^* = 0.35$ (from Figure 2). Remaining parameters are from the top panel of Figure 4. The solid line assumes no uncertainty about the long-run mean($\Omega = 0$). The dashed line is when DM learns about the long run ($\Omega > 1$).

4 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 demonstrate how expectations formed according to the framework introduced in sections 1 and 3, can lead to increased inflation variability. This variability worsens the central bank's policy trade-off between stabilizing inflation and output.

4.1 Firms' Optimal Price Setting

Suppose firm i reconsiders its price $P_{i,t}$ in period t. To maximize expected profits, it chooses a new price that will remain fixed until the next opportunity for adjustment. The firm's problem can be expressed as follows:[14]

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

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, departing from the conventional New Keynesian model's assumption of the full-information rational expectations. I propose that the firm's expectations are formed according to the cognitive limitations proposed in earlier sections, as will be discussed later.

The firm's optimal price $P^*_{i,t}$ is derived from its first-order condition. Expressed as a Taylor expansion around the zero-inflation steady state, this condition becomes

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

where I use lowercase to denote the log of the variable. Here, mc_{t+h} is the log of real

14. The new price maximizes the expected value of the firm's (current market value) profits. Since this pricing decision does not constrain any future decisions, 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.

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 D, the marginal costs do not depend on the quantity that a firm supplies because of the assumed feature of the production function (i.e., 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.[15] Let us define z_t as below to capture the aggregate terms in the firm's first-order condition:

$$z_t \equiv (1-\alpha\beta)(mc_t - mc) + \pi_t. \tag{4.1}$$

Thus, the firm's expectations of the current and future z_{t+h} pin down its subjectively optimal price:

$$p^*_{i,t} - p_{i,t-1} = E_{i,t}\left[\sum_{h=0}^{\infty} (\alpha\beta)^h z_{t+h}\right] \tag{4.2}$$

Since firm's objective depends only on aggregate conditions at the various dates $t+h$, 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.

4.2 Aggregate Economy

Real marginal costs. Real marginal costs are determined by the rest of the economy. As detailed in Appendix D, the household optimization problem and market-clearing conditions imply that

$$mc_t - mc = \chi\, x_t + e_t. \tag{4.3}$$

where χ is the sum of the coefficient of Relative Risk Aversion (of the utility function of consumption) and the Frisch elasticity of labor supply, and x_t is defined as $y_t - y^e_t$, where y^e_t is the efficient level of output. Finally, e_t is the cost-push shock[16] that is an i.i.d. draw from a zero mean Gaussian distribution.

Monetary policy. Given the existence of the cost-push shocks, the central bank cannot perfectly stabilize both inflation and the output. Thus, the central bank faces a policy

15. 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).

16. I do not take a stance on the source of cost-push shocks, but one example can be a time-varying, exogenous wage markup.

trade-off in stabilizing the two variables. To capture the trade-off succinctly, I assume that monetary policy is specified by a targeting rule of the form

$$x_t = -\frac{\theta}{1-\theta}\,\pi_t, \tag{4.4}$$

where 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 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.[17]

Aggregation. Price changes by individual firms aggregate into the overall inflation rate. The aggregate price index can be expressed as

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

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 $\pi_t = (1-\alpha)\,(p_t^* - p_{t-1})$. Therefore, we can derive the aggregate inflation by averaging the expectations of firms:

$$\pi_t = (1-\alpha)\bar{E}_t\left[\sum_{h=0}^{\infty}(\alpha\beta)^h\,z_{t+h}\right] \tag{4.5}$$

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

Determination of z_t. By substituting equations (4.3) and (4.4) into (4.1), we can derive that z_t is determined as

$$z_t = \left\{1 - (1-\alpha\beta)\,\chi\,\frac{\theta}{1-\theta}\right\}\pi_t + (1-\alpha\beta)\,e_t \tag{4.6}$$

Equations (4.5) and (4.6) 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.

17. The implication of such a targeting rule for the path of interest rates can be derived using the household inter-temporal optimization condition.

4.3 Firms' Macroeconomic Expectations

I assume firms form forecasts under the belief that z_t is an i.i.d. process as below.

$$z_t \sim \mathcal{N}\left(\mu,\, \sigma_z^2\right). \tag{4.7}$$

This assumption will align with FIRE as will be discussed. Importantly, firms may not have perfect awareness of the current value of z_t or the mean of the distribution from which it is drawn. (For simplicity, σ_z^2 is assumed to be known to DM.) Firms' prior beliefs about the mean are modeled as:

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

for some positive Ω. This is the environment similar to the one discussed in the previous section.

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:

$$\hat{z}_t = \lambda\left(1-\kappa_z\right)\hat{\mu}_{t-1} + \kappa_z\, z_t \tag{4.8}$$

The average expectation about the mean is

$$\hat{\mu}_t = \lambda\left(1-\kappa_\mu\right)\hat{\mu}_{t-1} + \kappa_\mu\, z_t. \tag{4.9}$$

where κ_z ad κ_μ are the long-run Kalman gains when updating beliefs about z_t and μ, respectively. Firms' beliefs are influenced by the realized z_t, which are determined by the rest of the aggregate economy, including the monetary policy.

4.4 Expectation Formations and Inflation Dynamics

Aggregate inflation is determined by firms' expectations about current and future values of z_t, as equation (4.5) illustrates. As detailed in equations (4.8) and (4.9), these expectations depend on 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 cost-push shock e_t determine the evolution of z_t, as described in equation (4.6). Combining these equations, we can infer that the inflation process is a linear function of e_t and $\hat{\mu}_{t-1}$:

$$\pi_t = \varphi_e\, e_t + \varphi_\mu\, \hat{\mu}_{t-1} \tag{4.10}$$

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

$$\varphi_e = \frac{\delta}{1 + \delta\,\chi\,\frac{\theta}{1-\theta} + \frac{1}{\alpha}\,\frac{1-\hat{\kappa}}{\hat{\kappa}}}$$
$$\varphi_\mu = \frac{1}{1 + \delta\,\chi\,\frac{\theta}{1-\theta} + \frac{1}{\alpha}\,\frac{1-\hat{\kappa}}{\hat{\kappa}}}\,\frac{1-\alpha}{\alpha}\,\frac{\hat{b}}{\hat{\kappa}},$$

where $\delta \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$, $\hat{\kappa} = \kappa_z + \kappa_\mu$, and $\hat{b} = \lambda(1-\kappa_z) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)$. Note that under perfect memory assumption, $\hat{\mu}_{t-1}$ is equal to the actual value of μ, which is set as zero. The derived φ_e and φ_μ converge to zeros as the central bank adopts full inflation targeting ($\theta \to 1$). See Appendix D 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 have learned the mean of its distribution. Therefore, firms expect the future marginal costs to be zero on average (since e_t fluctuates around zero) and set their prices to match the current marginal costs. Aggregate inflation is thus proportional to the realized z_t. The inflation process is derived as follows:

$$\pi_t = \frac{\delta}{1 + \delta\,\chi\,\frac{\theta}{1-\theta}}\,e_t$$

With conventional models of information frictions, firms are imperfectly aware of z_t though they have come to learn the true μ, which is zero in our example. Thus, their subjectively optimal price will still equal their perceived value of the current marginal costs (as they correctly expect that their future marginal costs to be zero on average). The inflation process is derived as

$$\pi_t = \frac{\delta}{1 + \delta\,\chi\,\frac{\theta}{1-\theta} + \frac{1}{\alpha}\,\frac{1-\kappa_z^*}{\kappa_z^*}}e_t,$$

where κ_z^* refers to the Kalman gain (when updating firms' belief about z_t) under the perfect-memory assumption. Intuitively, firms' reset prices are less responsive to the realized cost-push shocks than under FIRE because firms are not perfectly aware of z_t when resetting prices. Accordingly, aggregate inflation remains proportional to cost-push shocks, but the dependence is more muted compared to FIRE.

Under the proposed model, the inflation process is quite different from the above cases.

It 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 due to the fluctuating beliefs about the long run, as the coefficient ρ_μ is the serial correlation of $\hat{\mu}_t$.

4.5 Parameterization

I parameterize the model using standard values adopted in the literature. I set $\chi = 2$ to reflect that both the consumption utility and labor disutility functions are characterized as a log function. The time discount factor is set as $\beta = 0.99$ (the time unit is quarter). The frequency of price changes is chosen to match 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. The parameters describing the expectation process come from the estimation section. I take the mean estimates, which is $\phi_n = 0.36$, $\phi_m = 0.10$, and $\Omega = 0.32 \times Var[z_t]$. For detailed discussions, refer to the following section 5.

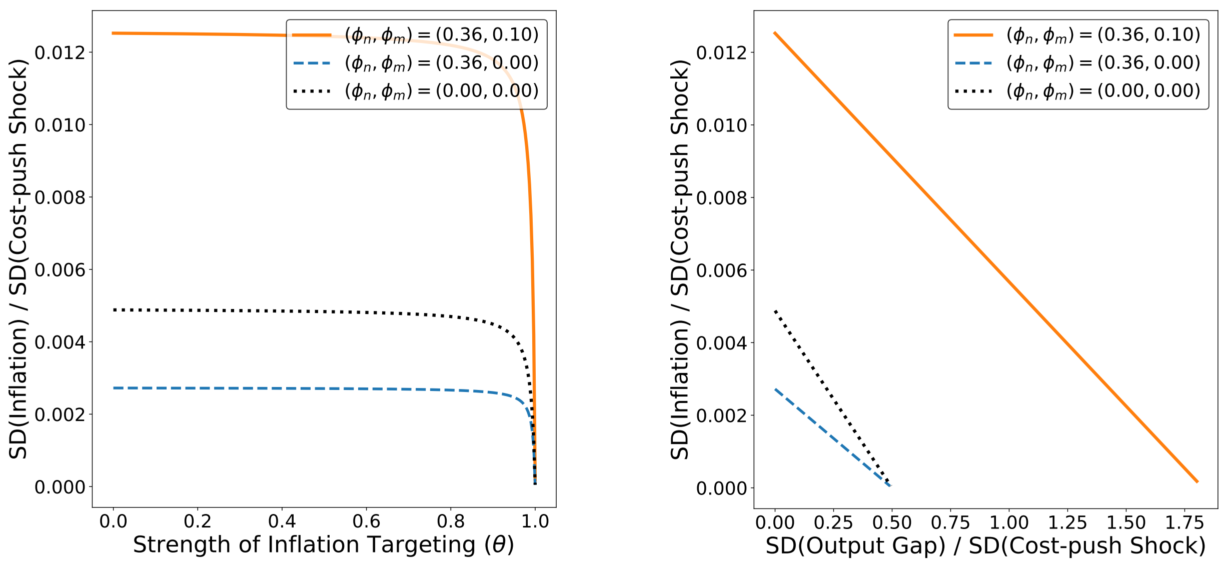
4.6 Monetary Policy and Inflation Variability

We have seen that the expectation-formation process shapes the statistical properties of inflation dynamics. In this section, I consider the strength of inflation targeting and its effect on inflation variability. The top left panel in Figure 6 shows the inflation variability for a given monetary policy rule on the x axis, which corresponds to the strength of the inflation targeting policy. I discuss the prediction for three different expectation assumptions: FIRE (black dotted line), conventional information friction model (blue dashed line), and proposed model (orange solid line). Naturally, stronger inflation targeting stabilizes the inflation process for all expectation assumptions. Furthermore, conventional information-friction models predict more stable inflation than under FIRE; since firms are not perfectly aware of the realized marginal cost, the prices do not correctly reflect firms' costs. In contrast, when firms are subject to both noisy news and noisy memory, they are imperfectly aware of the realized marginal cost and its long-run mean. Therefore, their expectations of future marginal costs fluctuate, inducing more price fluctuations. The top right panel of Figure 6 displays 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, indicating that for any output variability, the economy bears more variable inflation. This result highlights the likely challenge central bank may face when economic agents face the cognitive limitations of the sort discussed in the previous section.

Figure 6: The effect of monetary policy

(a) Inflation variability

(b) Stabilization trade-off



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; solid orange line correspond to the proposed model, blue dashed line to conventional information frictions models, and the black dotted line to the no-cognitive-noise benchmark. For each targeting rule θ on the x axis, the left panel displays the inflation variability, and the right panel shows the policy trade-off between stabilizing inflation and output gap. The model parameters are stated in the main text.

5 Estimation: The Extent of Cognitive Constraints

In this section, I estimate the two cognitive constraints using professional forecasters' survey data. I show that the conventional models of information frictions underestimate the severity of frictions.

5.1 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 participate in this survey. The earliest survey started in 1968, and I use survey forecasts made until the second quarter of 2022. For the estimation exercise, I focus on the Gross Domestic Output, both real and nominal, to investigate how forecasters perceive the overall economic activity in the U.S. The model environment illustrated in Section 1 and 3 is particularly suitable to describe the GDP forecasts in that forecasters can acquire information about the current state but not much about the long-term state.[18]

For data on the time series of macroeconomic variables, I use the Real-Time Data Set from the Federal Reserve Bank of Philadelphia. This data set provides the history of data releases for each variable, which can be particularly important in studying the variables in the National Income and Product Accounts (NIPA); they are often redefined or reclassified, due to which often the most recently available data may not correspond to the same variables forecast by the professional forecasters in the data set. The earliest release ("Advance" estimates) becomes available about a month after the quarter ends. Later releases of the NIPA incorporate more complete and detailed information. Thus, I treat the third estimates, released about three months after the quarter ends, as a true realization of the GDP.

5.2 Estimation Steps

Data-generating process. The unit of the forecasted variable I focus on is percent changes in GDP over twelve months. I furthermore suppose that such percent changes are described by the data-generating process in Section 1 and 3. Thus, each series has three components: trend, cycle, and irregular. I use an approximate Bandpass filter to decompose the third release of each GDP variable, following Baxter and King (1999).[19] The autoregressive process of the trend and cycle is then estimated using OLS. The results are reported in Table 1.

Estimation targets. The regression coefficients of interest are the slope terms of the specifications (2.6) and (2.5). For the estimation of regression (2.6), I use its "close-cousin" version to bypass several theoretical and statistical issues of the original specification proposed in

18. In comparison, long-term forecasts of inflation and interest rates are more likely to be influenced by the monetary authorities' policies, which requires careful modeling of all the term-structure related information available to forecasters, which is beyond the scope of the current paper.

19. The cutoff period for the trend component is set as 32 quarters and that for the irregular component is set as 3 quarters. 12 quarters of lags and leads each are included in the filter.

Coibion and Gorodnichenko (2015).[20] Based on the proposal of Goldstein (2023) and Gemmi and Valchev (2023), I use the following specification to measure how sticky forecasts are.[21]

$$F_{i,t}\, y_{t+h} - F_t\, y_{t+h} = \alpha_C + \beta_C\,(F_{i,t-1}\, y_{t+h} - F_{t-1}\, y_{t+h}) + error_{i,t+h|t-1} \tag{5.1}$$

This specification intuitively quantifies the size of Kalman gains when forecasters revise their views in response to news: If forecasters' relative positions are very persistent, their views are sticky and Kalman gains are small. Compared to Coibion and Gorodnichenko (2015), this specification partials out the effects of the common noise ($\tilde{\nu}_t$ in the current paper) by de-meaning individual forecasts.

Thus, two regression coefficients, β_C from equation (5.1) and β_I from equation (2.5), are used as estimation targets. For both GDP measures, three-quarter-ahead forecasts are used. I pool all forecasters to estimate the two regressions to overcome the power issue. Notably, pooling the data can be problematic if the relationships captured in equations (5.1) and (2.5) are different across forecasters. To limit such problems, I include forecaster-specific intercepts and cluster the standard errors two-way to account for correlations of the residuals within each forecaster and each period. Furthermore, I only use survey data from forecasters who participated long enough so that the time-variations within each individual are better reflected; specifically, I limit data to those with at least ten observations. Finally, I use Huber robust estimator to limit the influence of outliers. Table 2 reports the estimated regression coefficients.

Estimation strategy. I estimate two cognition parameters, ϕ_n and ϕ_m, that quantify the severity of the cognitive limitations in the attention and memory system. I pin down the two remaining parameters as follows. The noise in the external data, as described in equation (1.5), is approximated by the revision of the NIPA series. Since the gap between the third and the initial release largely reflects the limited real-time information availability, I use the standard deviation of such gap to pin down σ_ν in equation (1.13). The results are reported in the last panel of Table 1. Furthermore, I assume that the longest forecast horizon of the loss function (1.3) is eight quarters ahead since the SPF asks forecasters to submit their quarterly forecasts up to two years ahead, but the model prediction changes very little even for extremely longer forecast horizons.

20. As recognized by the authors, the regression coefficient is attenuated when noise in the external data is correlated among forecasters. The estimated coefficients are shown to be less reliable in other dimensions as well. Hajdini and Kurmann (2023) documents the fallibility of the estimation when structural changes are a concern, and Bianchi et al. (2022) documents the lack of out-of-sample predictability. These findings suggest a possibility of a small sample problem; the time-series estimation is likely too noisy to test the statistical relationship, given the modest length of the data series.

21. I keep the notation β_C to reduce notation burdens.

Table 1: Data-generating Process

	ρ	σ_ϵ	ρ_μ	σ_μ	σ_η	σ_ν
Real GDP	0.878	1.098	0.991	0.105	0.369	0.143
Nomianl GDP	0.872	1.089	0.999	0.112	0.409	0.136

The first two panels of the table report the estimated data-generating process of both GDP measures. Approximate Bandpass filter is used to decompose the series into trend, cycle, and irregular components. The exact details of the filter follow the discussion in Baxter and King (1999). An AR(1) model is used to approximate the process of the cycle (ρ and σ_ϵ) and the trend (ρ_μ and σ_μ), both estimated using OLS. The standard deviation of the irregular component (σ_η) is listed in the third panel. The last panel describes the noise in the external data (σ_ν), as approximated by the revision of the NIPA series.

Table 2: Estimation Targets: Regression Coefficients

	β_C	SD	p-value	β_I	SD	p-value
Real GDP	0.54	0.01	0.00	-0.17	0.05	0.00
Nominal GDP	0.54	0.02	0.00	-0.23	0.05	0.00

This table reports the regression coefficients used as estimation targets for each GDP measure. The first panel displays the coefficient estimated in regression (5.1). The second panel shows the regression coefficient estimates of (2.5). For both, I report the results of Huber robust estimator when individual fixed effects are included, and the standard errors are two-way clustered in each forecaster and date.

5.3 Estimation Results

The first panel of Table 3 reports the estimates of noisy news and noisy memory (ϕ_n and ϕ_m) and their standard errors. These estimates almost exactly match the estimation targets in Table 2 (though it is not guaranteed to do so), and the standard errors are quite tight as well. The second panel shows the estimated size and standard errors of ϕ_n when assuming perfect memory. This result corresponds to the estimation exercises proposed using conventional information friction models. Estimated ϕ_n, assuming both noisy news and noisy memory, is about 50% larger than those assuming noisy news alone. This point is made in Section 2, where I showed that the methodology in Coibion and Gorodnichenko (2015) underestimates the magnitude of ϕ_n because it mis-attributes the extra sensitivity from noisy memory to low ϕ_n.

Table 4 assesses the model's capability in explaining non-targeted moments. I investigate whether the estimated model can generate realistic variations observed in survey data. For this purpose, 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). The data moments are estimated using Huber Robust estimator, and individual fixed effects are controlled when estimating the cross-section moments. I compare these empirical moments to the model predictions of the same objects. The estimated model provides a reasonable quantitative fit for most variations.

Table 3: Estimation Results: Cognitive Parameters

	noisy news & noisy memory				noisy news only			
	ϕ_n	SD	ϕ_m	SD	ϕ_n	SD	ϕ_m	SD
Real GDP	0.43	0.07	0.12	0.04	0.29	0.02	0.0	-
Nominal GDP	0.28	0.06	0.08	0.03	0.18	0.01	0.0	-

This table reports the estimated severity of cognitive limitations, "noisy news" (ϕ_n) and "noisy memory" (ϕ_m). The first panel displays the estimated size of ϕ_n and ϕ_m and their standard errors. The second panel shows the estimation under the perfect-memory assumption; the estimated size of ϕ_n and the standard error is reported.

Table 4: Estimation Fit: Variations

	Variations in Forecasts				Variations in Revisions			
	Cross-section		Time		Cross-section		Time	
	Data	Model	Data	Model	Data	Model	Data	Model
Real GDP	0.60	0.54	1.11	1.52	0.53	0.51	0.48	0.13
Nominal GDP	0.76	0.93	2.23	5.51	0.70	0.87	0.51	0.20

The table compares the predictions of the estimated model with the non-targeted data moments. The first panel reports variations in forecasts in cross-sections and time. The second panel shows variations in forecast revisions in cross-sections and time. The data moments are estimated using Huber Robust estimator; for cross-section moments, forecaster fixed effects are controlled.

6 Conclusion

This study aims to identify the constraints that prevent economic agents from developing expectations aligned with full information rational expectations. My analysis demonstrates that the finite capacity to process information, both internal and external, explains the diverse characteristics of survey forecasts that traditional expectations-formation models have difficulty addressing.

Recognizing these cognitive limitations allows us to gain valuable insights into longer-term expectation, where data is often scarce. My analysis predicts that long-run expectations can exhibit overreactions, even when the near-term expectations are revised sluggishly. This insight carries far-reaching implications for understanding how economic agents perceive the long-run economy. It suggests that even seemingly stable long-run expectations may become unanchored. In particular, using an illustrative macroeconomic model, I show that the cognitive limitations discussed in this paper can make the central bank's trade-off in stabilizing both inflation and the output gap more challenging.

The prediction of horizon-dependent forecast sensitivity, which is not immediately obvious ex ante, highlights the need for better models of how economic agents form macroeconomic perceptions. This study examines the cognitive constraints influencing these perceptions, provides a tractable model to analyze and estimate them, and demonstrates how integrating empirically validated expectations-formation models can enhance macroeconomic analysis.

References

Afrouzi, Hassan, Spencer Y. Kwon, Augustin Landier, Yueran Ma, and David Thesmar. 2023. "Overreaction in Expectations: Evidence and Theory." *Quarterly Journal of Economics* 138 (3).

Afrouzi, Hassan, and Choongryul Yang. 2021. "Dynamic Rational Inattention and the Phillips Curve."

Andrade, Philippe, and Hervé Le Bihan. 2013. "Inattentive professional forecasters." *Journal of Monetary Economics* 60 (8): 967–982.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2020. "Business-Cycle Anatomy." *American Economic Review* 110 (10): 3030–3070.

Angeletos, George-Marios, and Zhen Huo. 2021. "Myopia and Anchoring." *American Economic Review* 111 (4): 1166–1200.

Angeletos, George-Marios, Zhen Huo, and Karthik A. Sastry. 2021. "Imperfect Macroeconomic Expectations: Evidence and Theory." *NBER Macroeconomics Annual* 35 (1): 2–86.

Baxter, Marianne, and Robert G. King. 1999. "Measuring business cycles: Approximate band-pass filters for economic time series." *Review of Economics and Statistics* 81 (4).

Bianchi, Francesco, Sydney C. Ludvigson, and Sai Ma. 2022. "Belief Distortions and Macroeconomic Fluctuations." *American Economic Review* 112 (7): 2269–2315.

Bordalo, Pedro, John J. Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer. 2022. "Memory and Probability." *The Quarterly Journal of Economics* 138 (1): 265–311.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Overreaction in Macroeconomic Expectations." *American Economic Review* 110 (9): 2748–2782.

Bordalo, Pedro, Nicola Gennaioli, Rafael La Porta, and Andrei Shleifer. 2023. "Belief Overreaction and Stock Market Puzzles." *Journal of Political Economy.*

Bordalo, Pedro, Nicola Gennaioli, Rafael R.L. Porta, and Andrei Shleifer. 2019. "Diagnostic Expectations and Stock Returns." *The Journal of Finance* 74 (6): 2839–2874.

Broer, Tobias, and Alexandre N. Kohlhas. 2022. "Forecaster (Mis-)Behavior." *The Review of Economics and Statistics,* 1–45.

Carvalho, Carlos, Stefano Eusepi, Emanuel Moench, and Bruce Preston. 2023. "Anchored Inflation Expectations." *American Economic Journal: Macroeconomics* 15 (1): 1–47.

Case, Karl E, Robert J Shiller, and Anne K Thompson. 2012. "What Have They Been Thinking? Homebuyer Behavior in Hot and Cold Markets." *Brookings Papers on Economic Activity265* 43 (2): 265–315.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *American Economic Review* 105 (8): 2644–2678.

———. 2012. "What can survey forecasts tell us about information rigidities?" *Journal of Political Economy* 120 (1): 116–159.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *American Economic Review* 108 (9): 2671–2713.

Coibion, Olivier, Yuriy Gorodnichenko, and Michael Weber. 2022. "Monetary Policy Communications and Their Effects on Household Inflation Expectations." *Journal of Political Economy* 130 (6): 1537–1584.

Collin-Dufresne, Pierre, Michael Johannes, and Lars A. Lochstoer. 2016. "Parameter Learning in General Equilibrium: The Asset Pricing Implications." *American Economic Review* 106 (3): 664–98.

Cover, Thomas M., and Joy A. Thomas. 2005. *Elements of Information Theory.*

Crump, Richard K., Stefano Eusepi, Emanuel Moench, and Bruce Preston. 2023. "The term structure of expectations." *Handbook of Economic Expectations,* 507–540.

d'Arienzo, Daniele. 2020. "Maturity Increasing Overreaction and Bond Market Puzzles." *SSRN Electronic Journal.*

Elliott, Graham, Ivana Komunjer, and Allan Timmermann. 2008. "Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss?" *Journal of the European Economic Association* 6 (1): 122–157.

Enke, Benjamin, and Thomas Graeber. 2023. "Cognitive Uncertainty." *The Quarterly Journal of Economics* 138 (4): 2021–2067.

Evans, George W., and Seppo Honkapohja. 2012. *Learning and expectations in macroeconomics.*

Farmer, Leland, Emi Nakamura, and Jon Steinsson. 2024. "Learning About the Long Run." *Journal of Political Economy.*

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. "Natural Expectations and Macroeconomic Fluctuations." *Journal of Economic Perspectives* 24 (4): 67–84.

Galí, Jordi. 2015. "Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition." *Economics Books.*

Gáti, Laura. 2023. "Monetary Policy & Anchored Expectations: An Endogenous Gain Learning Model." *Journal of Monetary Economics* 140:S37–S47.

Gemmi, Luca, and Rosen Valchev. 2023. "Biased Surveys." *National Bureau of Economic Research* Working Paper Series (31607).

Giglio, Stefano, and Bryan Kelly. 2018. "Excess Volatility: Beyond Discount Rates." *The Quarterly Journal of Economics* 133 (1): 71–127.

Goldstein, Nathan. 2023. "Tracking Inattention." *Journal of the European Economic Association* 21 (6): 2682–2725.

Hajdini, Ina, and André Kurmann. 2023. "Predictable Forecast Errors in Full-Information Rational Expectations Models with Regime Shifts."

Hamilton, James D. 2018. "Why You Should Never Use the Hodrick-Prescott Filter." *The Review of Economics and Statistics* 100 (5): 831–843.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson. 2022. "The Slope of the Phillips Curve: Evidence from U.S. States." *The Quarterly Journal of Economics* 137 (3): 1299–1344.

Juodis, Artūras, and Simas Kučinskas. 2023. "Quantifying noise in survey expectations." *Quantitative Economics* 14 (2): 609–650.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A Rational Theory of Mutual Funds' Attention Allocation." *Econometrica* 84 (2): 571–626.

Kohlhas, Alexandre N., and Ansgar Walther. 2021. "Asymmetric Attention." *American Economic Review* 111 (9): 2879–2925.

Kučinskas, Simas, and Florian S Peters. 2022. "Measuring Under- and Overreaction in Expectation Formation." *The Review of Economics and Statistics,* 1–45.

Maćkowiak, Bartosz, and Mirko Wiederholt. 2009. "Optimal Sticky Prices under Rational Inattention." *American Economic Review* 99 (3): 769–803.

Malmendier, Ulrike, and Stefan Nagel. 2016. "Learning from inflation experiences." *Quarterly Journal of Economics* 131 (1).

Malmendier, Ulrike, and Jessica A. Wachter. 2021. "Memory of Past Experiences and Economic Decisions." *SSRN Electronic Journal.*

Mankiw, N. Gregory, and Ricardo Reis. 2002. "Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve." *The Quarterly Journal of Economics* 117 (4): 1295–1328.

Miao, Jianjun, Jieran Wu, and Eric R Young. 2022. "Multivariate Rational Inattention." *Econometrica* 90 (2): 907–945.

Nagel, Stefan, and Zhengyang Xu. 2022. "Asset Pricing with Fading Memory." *The Review of Financial Studies* 35 (5): 2190–2245.

Nordhaus, William D. 1987. "Forecasting Efficiency: Concepts and Applications." *The Review of Economics and Statistics* 69 (4): 667.

Orphanides, Athanasios, and John C. Williams. 2006. "Monetary Policy with Imperfect Knowledge." *Journal of the European Economic Association* 4 (2-3): 366–375.

Paivio, Allan. 1990. *Mental representations: A dual coding approach.* Oxford university press. ISBN: 0195362004.

Patton, A. J., and Allan Timmermann. 2010. "Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion." *Journal of Monetary Economics* 57 (7): 803–820.

Rotemberg, Julio J., and Michael Woodford. 1997. "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy." *NBER Macroeconomics Annual* 12:284–346.

Shiller, Robert J., and Anne K. Thompson. 2022. "What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets: A Ten-Year Retrospect." *Brookings Papers on Economic Activity* 2022-Spring.

Silveira, Rava Azeredo da, Yeji Sung, and Michael Woodford. 2020. “Optimally Imprecise Memory and Biased Forecasts.” *National Bureau of Economic Research* Working Paper Series (28075).

Sims, Christopher A. 2003. “Implications of rational inattention.” *Journal of Monetary Economics* 50 (3).

Singleton, Kenneth J. 2021. “Presidential Address: How Much “Rationality” Is There in Bond-Market Risk Premiums?” *The Journal of Finance* 76 (4): 1611–1654.

Timmermann, Allan G. 1993. “How learning in financial markets generates excess volatility and predictability in stock prices.” *Quarterly Journal of Economics* 108 (4): 1135–1145.

Wachter, Jessica A., and Michael J. Kahana. 2023. “Associative Learning and Representativeness.” *SSRN Electronic Journal.*

Wachter, Jessica A., and Michael Jacob Kahana. 2024. “A Retrieved-Context Theory of Financial Decisions.” *The Quarterly Journal of Economics* 139 (2): 1095–1147.

Wang, Chen. 2021. “Under- and Overreaction in Yield Curve Expectations.” *SSRN Electronic Journal.*

Weitzman, Martin L. 2007. “Subjective Expectations and Asset-Return Puzzles.” *American Economic Review* 97 (4): 1102–1130.

Woodford, Michael. 2003. “Imperfect Common Knowledge and the Effects of Monetary Policy.” In *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps,* edited by Michael Woodford, Philippe Aghion, Roman Frydman, and Joseph Stiglitz, 25–58. Princeton University Press.

———. 2020. “Modeling Imprecision in Perception, Valuation, and Choice.” *Annual Review of Economics* 12 (1): 579–601.

Zarnowitz, Victor. 1985. “Rational Expectations and Macroeconomic Forecasts.” *Journal of Business & Economic Statistics* 3 (4): 293.

For Online Publication
APPENDIX

Sung,
"Macroeconomic Expectations and Cognitive Noise"

A Derivation of the Optimal Cognitive Process

The law of motion of the state vector x_t is described as below.

$$x_t = d + A\, x_{t-1} + B\, \epsilon_t, \quad \epsilon_t \sim \mathcal{N}\left(O,\, \Sigma_\epsilon\right)$$

where d, A, B, Σ_ϵ are constant parameters *known* to DM. (When analyzing the assumption of the unknown μ, I put μ inside x_t.) And DM's "default" prior about x_t is described as

$$x_t \sim \mathcal{N}\left(\mu_x,\, \Sigma_x\right)$$

Below I discuss the statistical properties of DM's forecasts under the probability measures that jointly consider the state vector x_t, DM's cognitive states, $m_{i,t}$ and $n_{i,t}$, and the DM's default prior about x_t. Thus, these represent DM's *perceived* probability distribution, rather than the '*actual* distribution.

For any given state vector x_t, I show the optimal structure of the cognitive process, described by the sequence of $\{K_t, \Sigma_{ut}, \Lambda_t, \Sigma_{\omega t}\}_{t=0}^{\infty}$, that minimizes the loss function (1.11) subject to the information environment (1.6), (1.7), (1.9), and (1.10).

A.1 The Optimal Structure for the Representation

I show below the optimal encoding of the cognitive state, $m_{i,t}$ and $n_{i,t}$. In particular, I show that the optimal $m_{i,t}$ stores $E\left[x_t \middle| M_{i,t}\right]$ with noise while the optimal $n_{i,t}$ records $E\left[x_t \middle| N_t\right]$ with noise. Thus, the dimension of the optimal $m_{i,t}$ and $n_{i,t}$ is no bigger than the dimension of the state vector x_t.

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

Partition of $m_{i,t}$ We can partition $m_{i,t} = \Lambda_t \cdot M_{i,t} + \omega_{i,t}$ in the following form

$$\begin{pmatrix} \vec{m}_{i,t} \\ \tilde{m}_{i,t} \end{pmatrix} = \begin{pmatrix} \Lambda_{at} & \Lambda_{b,t} \\ \Lambda_{ct} & \Lambda_{dt} \end{pmatrix} \begin{pmatrix} \vec{M}_{i,t-1} \\ E\left[x_t \middle| M_{i,t}\right] \end{pmatrix} + \begin{pmatrix} \vec{\omega}_{i,t} \\ \tilde{\omega}_{i,t} \end{pmatrix} \tag{A.1}$$

where the elements of $\vec{M}_{i,t}$ are orthogonal to $E\left[x_t \middle| M_{i,t}\right]$ and $\vec{M}_{i,t-1}$ and $E\left[x_t \middle| M_{i,t}\right]$ span the same vector space as $M_{i,t}$. The terms $\vec{\omega}_{i,t}$ and $\tilde{\omega}_{i,t}$ are Gaussian innovations. I impose the following normalization assumption

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

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

$$Cov\left[E[x_t|\, M_{i,t}] - \tilde{m}_{i,t}, \vec{m}_{i,t}\right] = \vec{O} \tag{A.2a}$$
$$Cov\left[E[x_t|\, M_{i,t}] - \tilde{m}_{i,t}, \tilde{m}_{i,t}\right] = O \tag{A.2b}$$

The second requirement implies that

$$\begin{aligned} & Cov\left[E[x_t|\, M_{i,t}], \tilde{m}_{i,t}\right] = Var\left[\tilde{m}_{i,t}\right] \\ \Leftrightarrow \quad & Var\left[\Lambda_{ct}\, \vec{M}_{i,t} + \tilde{\omega}_{i,t}\right] = (1 - \Lambda_{dt})\, Var\left[E[x_t|\, M_{i,t}]\right]\, \Lambda'_{dt} \end{aligned} \tag{A.3}$$

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

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

$$\begin{pmatrix} \vec{n}_{i,t} \\ \tilde{n}_{i,t} \end{pmatrix} = \begin{pmatrix} K_{at} & K_{bt} \\ K_{ct} & K_{dt} \end{pmatrix} \begin{pmatrix} \vec{N}_t \\ E[x_t|\, N_t] \end{pmatrix} + \begin{pmatrix} \vec{u}_{i,t+1} \\ \tilde{u}_{i,t+1} \end{pmatrix} \tag{A.4}$$

where $\vec{N}_t$ is the components in N_t that are orthogonal to $E[x_t|\, N_t]$, thus $E[x_t|\, N_t]$ and that $\vec{N}_t$ and $E[x_t|\, N_t]$ span the same vector space as N_t. Both $\vec{u}_{i,t+1}$ and $\tilde{u}_{i,t+1}$ are Gaussian innovations. As before, I impose the following normalization assumption

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

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

$$Cov\left[x_t - \tilde{n}_{i,t}, \vec{n}_{i,t}|\, m_{i,t}\right] = \vec{O} \tag{A.5a}$$
$$Cov\left[x_t - \tilde{n}_{i,t}, \tilde{n}_{i,t}|\, m_{i,t}\right] = O \tag{A.5b}$$

We can see that (A.5b) implies

$$\begin{aligned} & Cov\left[x_t, K_{dt}\, E[x_t|\, N_t]|\, m_{i,t}\right] = Var\left[K_{ct}\, \vec{N}_t + K_{dt}\, E[x_t|\, N_t] + \tilde{u}_{i,t} \,\middle|\, m_{i,t}\right] \\ \Leftrightarrow \quad & Var\left[K_{ct}\, \vec{N}_t + \tilde{u}_{i,t} \,\middle|\, m_{i,t}\right] = Cov\left[x_t, K_{dt}\, E[x_t|\, N_t]|\, m_{i,t}\right] - K_{dt}\, Var\left[E[x_t|\, N_t]|\, m_{i,t}\right]\, K'_{dt} \end{aligned} \tag{A.6}$$

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

Without loss of generality, I express that

$$E[x_t|\, N_t] = C\, x_t + \nu_t, \quad \nu_t \sim \mathcal{N}(O, \Sigma_\nu) \tag{A.7}$$

where the matrices C and Σ_ν are known to DM. This expression further simplifies the normalization condition (A.6).

$$Var\left[K_{ct}\,\vec{N}_t+\tilde{u}_{i,t}\middle|\,m_{i,t}\right]=(I-K_{dt}C)\,Var\left[x_t|\,m_{i,t}\right](K_{dt}C)'-K_{dt}\Sigma_\nu K_{dt}' \tag{A.8}$$

Step 2: Forecast accuracy depends only on K_{dt} and Λ_{dt}

We observe from (A.1) that

$$E\left[x_t|\,M_{i,t}\right]|\,m_{i,t}=E\left[x_t|\,M_{i,t}\right]|\,\tilde{m}_{i,t}$$

That is, the information in $m_{i,t}$ about $E\left[x_t|\,M_{i,t}\right]$ is completely captured by $\tilde{m}_{i,t}$, which follows from (A.2a). We can furthermore see that Λ_{dt} uniquely determines the prior uncertainty $E\left[x_t|\,m_{i,t}\right]$.

$$\begin{aligned}Var\left[x_t|\,\tilde{m}_{i,t}\right]&=Var\left[x_t\right]-Cov\left[\tilde{m}_{i,t},x_t\right]\\&=Var\left[x_t\right]-\Lambda_{dt}\,Cov\left[E\left[x_t|\,M_{i,t}\right],x_t\right]\\&=Var\left[x_t\right]-\Lambda_{dt}\,Var\left[E\left[x_t|\,M_{i,t}\right]\right]\end{aligned}$$

Likewise, we also observe from the proposed partition (A.4) 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 (A.5a). Furthermore, we can see that K_{dt} uniquely determines the posterior uncertainty.

$$\begin{aligned}Var\left[x_t|\,m_{i,t},\tilde{n}_{i,t}\right]&=Var\left[x_t|\,m_{i,t}\right]-Cov\left[\tilde{n}_{i,t},x_t|\,m_{i,t}\right]\\&=Var\left[x_t|\,m_{i,t}\right]-K_{dt}\,Cov\left[E\left[x_t|\,N_t\right],x_t|\,m_{i,t}\right]\\&=(I-K_{dt}C)\,Var\left[x_t|\,m_{i,t}\right]\end{aligned}$$

Thus, for a given level of $Var\left[x_t|\,M_{i,t}\right]$, which is predetermined at time t, the matrices Λ_{dt} and K_{dt} uniquely determine the variances $Var\left[x_t|\,\tilde{m}_{i,t}\right]$ and $Var\left[x_t|\,m_{i,t},\tilde{n}_{i,t}\right]$.

Step 3: The Optimal Choice of Λ_t and K_t

Since the elements of Λ_t and K_t other than Λ_{dt} and K_{dt} 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

$$I\left(m_{i,t};M_{i,t}\right)=I\left(\left(\vec{m}_{i,t},\tilde{m}_{i,t}\right);\left(\vec{M}_{i,t-1},E\left[x_t|\,M_{i,t}\right]\right)\right)$$

As discussed in Appendix C.2 of Silveira et al. (2020), the lower bound is equal to $I\left(\tilde{m}_{i,t};E\left[x_t|\,M_{i,t}\right]\right)$, which again is achieved when $\Lambda_{at}=\Lambda_{b,t}=\Lambda_{ct}=O$. Likewise,

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

whose lower bound of this mutual information is equal to $I\left(\tilde{n}_{i,t};E\left[x_t|\,N_t\right]\right)$. This lower bound is achieved when $K_{at}=K_{bt}=K_{ct}=O$.

Summary

Using the following notations to describe the DM's prior and posterior belief,

$$x_t |\, m_{i,t} \sim \mathcal{N}\left(x^m_{i,t|t},\, \Sigma^m_{t|t}\right) \tag{A.9}$$

$$x_t |\, m_{i,t}, n_{i,t} \sim \mathcal{N}\left(x_{i,t|t},\, \Sigma_{t|t}\right) \tag{A.10}$$

The optimal representation of internal information has the following law of motion.

$$\tilde{m}_{i,t} = \tilde{\Lambda}_t\, x^m_{i,t|t} + \tilde{\omega}_{i,t} \tag{A.11}$$

where the variance of memory noise $\tilde{\omega}_{i,t}$ is determined from (A.3) to be

$$\Sigma_{\omega t} = (I - \tilde{\Lambda}_t)\,(\Sigma_x - \Sigma_{t|t-1})\,\tilde{\Lambda}'_t \tag{A.12}$$

This representation yields the time-t prior belief as follows.

$$x^m_{i,t|t} = (I - \tilde{\Lambda}_t)\mu_x + \tilde{\Lambda}_t x_{i,t|t-1} + \tilde{\omega}_{i,t} \tag{A.13}$$

$$\Sigma^m_{t|t} = (I - \tilde{\Lambda}_t)\Sigma_x + \tilde{\Lambda}_t\,\Sigma_{t|t-1} \tag{A.14}$$

Likewise, the optimal representation of external information results in

$$\tilde{n}_{i,t} = \tilde{K}_t\, x_t + \tilde{\nu}_t + \tilde{u}_{i,t} \tag{A.15}$$

where the variance of attention noise $\tilde{u}_{i,t}$ is derived from (A.8) as

$$\Sigma_{ut} = (I - \tilde{K}_t)\Sigma^m_{t|t}\,\tilde{K}'_t - \tilde{\Sigma}_\nu \tag{A.16}$$

This representation characterizes the time-t posterior belief as

$$x_{i,t|t} = (I - \tilde{K}_t)\, x^m_{i,t|t} + \tilde{K}_t\, x_t + \tilde{\nu}_t + \tilde{u}_{i,t} \tag{A.17}$$

$$\Sigma_{t|t} = (I - \tilde{K}_t)\,\Sigma^m_{t|t} \tag{A.18}$$

where I redefine $\tilde{K}_t \equiv K_{dt}\, C$ and $\tilde{\nu}_t \equiv K_{dt}\,\nu_t$ (so that $\tilde{\Sigma}_\nu \equiv K_{dt}\Sigma_\nu K'_{dt}$).

A.2 Specification of $\tilde{K}_t$

Before deriving the feasible set of $\tilde{K}_t$, I make a simplifying assumption that the news vector provides information about some linear combination of the state vector x_t. In other words, we can express that

$$N_t = c'\, x_t + \bar{\nu}_t, \quad \bar{\nu}_t \sim \mathcal{N}\left(0,\, \sigma^2_\nu\right) \tag{A.19}$$

This also implies that we can express the optimal $n_{i,t}$ as

$$\tilde{n}_{i,t} = \kappa_t\, \bar{n}_{i,t} \tag{A.20}$$

where κ_t is a column vector and $\bar{n}_{i,t}$ is a uni-variate random variable defined as

$$\bar{n}_{i,t} = \kappa_{ct}\,(c' x_t + \bar{\nu}_t) + \bar{u}_{i,t}$$

for some positive scalar κ_{ct} that remains to be specified, and the idiosyncratic noise $\bar{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}(0,\, \sigma_{ut}^2)$. In summary, the noisy representation of external information is described as (A.15) where the loading matrix $\tilde{K}_t$, the variance of $\tilde{\nu}_t$, and the variance of attention noise $\tilde{u}_{i,t}$ are derived as a function of κ_t, κ_{ct}, and σ_{ut}^2 as follows.

$$\begin{aligned}
\tilde{K}_t &= \kappa_{ct}\, \kappa_t\, c' \\
\Sigma_\nu &= \sigma_\nu^2\, \kappa_{ct}^2\, \kappa_t\, \kappa_t' \\
\Sigma_{ut} &= \sigma_{ut}^2\, \kappa_t\, \kappa_t'
\end{aligned}$$

The normalization assumption (A.5b) implies that

$$\kappa_{ct}\, \Sigma_{t|t}^m\, c\, \kappa_t' = \left(\kappa_{ct}^2 (c' \Sigma_{t|t}^m\, c + \sigma_\nu^2) + \sigma_{ut}^2\right) \kappa_t\, \kappa_t'$$

With e_i denoting the column vector that picks out the ith element of x_t, it must then be that

$$e_i'\, \kappa_t = \frac{\kappa_{ct}\, (e_i'\, \Sigma_{t|t}^m\, c)}{\kappa_{ct}^2 (c' \Sigma_{t|t}^m\, c + \sigma_\nu^2) + \sigma_{ut}^2} \tag{A.21}$$

for all i. I further normalize the first element of the column vector κ_t to be one, i.e., $e_1'\, \kappa_t = 1$. This implies that

$$\sigma_{ut}^2 = \kappa_{ct} (e_1' \Sigma_{t|t}^m c) - \kappa_{ct}^2 (c' \Sigma_{t|t}^m c + \sigma_\nu^2). \tag{A.22}$$

Thus, the value of σ_{ut}^2 will be determined as a function of κ_{ct}, and any $\kappa_{ct} \in \left[0, \frac{e_1' \Sigma_{t|t}^m c}{c' \Sigma_{t|t}^m c + \sigma_\nu^2}\right]$ ensures that the resulting σ_{ut}^2 is non-negative. Substituting (A.22) into the (A.21) results in the expression for the column vector κ_t as

$$\kappa_t = \frac{1}{e_1'\, \Sigma_{t|t}^m\, c}\, \Sigma_{t|t}^m\, c. \tag{A.23}$$

Using the information constraint, we can derive that

$$\begin{aligned}
I\,(n_{i,t}; N_t) &= I\,(\bar{n}_{i,t}; N_t) \\
&= -\frac{1}{2} \log \left(1 - \frac{\kappa_{ct}^2\, Var\,[N_t]}{\kappa_{ct}^2\, Var\,[N_t] + \sigma_{ut}^2}\right)
\end{aligned}$$

The constraint $I\,(n_{i,t}; N_t) \leq -\frac{1}{2} \log\, \phi_n$ then implies that

$$\frac{\sigma_{ut}^2}{\kappa_{ct}^2} \geq \frac{\phi_n}{1 - \phi_n}\, Var\,[N_t]$$

Substituting (A.22) yields that

$$\kappa_{ct} \leq \frac{e_1' \Sigma_{t|t}^m c}{c' \Sigma_{t|t}^m c + \sigma_\nu^2 + \frac{\phi_n}{1-\phi_n} (c' \Sigma_x c + \sigma_\nu^2)} \tag{A.24}$$

Thus the optimal value of κ_{ct} corresponds to the upper bound in (B.34). Then, the resulting

σ_{ut}^2 is

$$\sigma_{ut}^2 = \kappa_{ct}^2 \left(\frac{\phi_n}{1-\phi_n}\right)\left(c'\Sigma_x c + \sigma_\nu^2\right) \tag{A.25}$$

A.3 Specification of $\tilde{\Lambda}_t$

The mutual-information capacity (1.10) constrains the choice of $\tilde{\Lambda}_t$. We can derive that

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

Therefore, it remains to specify $\tilde{\Lambda}_t$ that satisfies the following inequality.

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

Below I describe the optimization problem to pin down the optimal $\tilde{\Lambda}_t$. First, I discuss a chance of variable to ease the optimization.

The Change of the Choice Variable

There are two requirements for the feasibility of $\tilde{\Lambda}_t$. First, the resulting $\Sigma_{t|t}^m$ is a symmetric and positive semidefinite matrix. Second, the diagonal elements of $\Sigma_{t|t}^m$ 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 $\Sigma_{t|t}^m - \Sigma_{t|t-1}$ and $\Sigma_x - \Sigma_{t|t}^m$ 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 (1.10) 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.

The Constraints

The prior uncertainty is formed according to

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

Thus, we can see that the matrix W_t is not only positive semidefinite, but also symmetric. And the posterior uncertainty can be described as

$$\Sigma_{t|t} = (I - \tilde{K}_t)\, \Sigma^m_{t|t} = \Sigma^m_{t|t} - \frac{1}{\Omega^m_{t|t}}\, \Sigma^m_{t|t}\, c\, c'\, \Sigma^m_{t|t}$$

where $\Omega^m_{t|t}$ is defined as

$$\Omega^m_{t|t} \equiv c'\, \Sigma^m_{t|t}\, c + \sigma^2_\nu + \frac{\phi_n}{1-\phi_n}\left(c'\Sigma_x\, c + \sigma^2_\nu\right).$$

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

$$\begin{aligned}\Sigma^m_{t|t} - \Sigma_{t|t-1} &= \left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \; W_t \; \left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \\ \Omega^m_{t|t} &= c'\, \Sigma^m_{t|t}\, c + \sigma^2_\nu + \frac{\phi_n}{1-\phi_n}\left(c'\Sigma_x\, c + \sigma^2_\nu\right) \\ \Sigma_{t|t} &= \Sigma^m_{t|t} - \Sigma^m_{t|t}\, c\, \left(\Omega^m_{t|t}\right)^{-1} c'\, \Sigma^m_{t|t}\end{aligned}$$

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.

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 $\bar{\Upsilon}_2$. 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 $tr\,(\Upsilon_1\,W_t) \geq 0$ and $tr\,(\Upsilon_2\,(I - W_t)) \geq 0$. This is because $tr\,(\Upsilon_1\,W_t) = tr\left(\bar{\Upsilon}_1\,(I - D)\right)$ and $tr\,(\Upsilon_2\,(I - W_t)) = tr\left(\bar{\Upsilon}_2\,D\right)$.

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.

The Lagrangian Problem and the First Order Conditions

The Lagrangian problem is as follows.

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

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)

$$
\begin{aligned}
-\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}\Gamma_1\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} + \Upsilon_1 - \Upsilon_2 + \mu\,\det\left(W_t\right)W_t^{-1} &= O && \text{(A.26a)} \\
\Gamma_1 - c\,\Gamma_2\,c' - \Gamma_3 + c\left(\Omega^m_{t|t}\right)^{-1} c'\,\Sigma^m_{t|t}\,\Gamma_3 + \Gamma_3\,\Sigma^m_{t|t}\,c\left(\Omega^m_{t|t}\right)^{-1} c' &= O && \text{(A.26b)} \\
\Gamma_2 - \left(\Omega^m_{t|t}\right)^{-1} c'\,\Sigma^m_{t|t}\,\Gamma_3\,\Sigma^m_{t|t}\,c\,\left(\Omega^m_{t|t}\right)^{-1} &= O && \text{(A.26c)} \\
-Q + \Gamma_3 &= O && \text{(A.26d)}
\end{aligned}
$$

and the slackness conditions are

$$
\begin{aligned}
\Upsilon_1 W_t = O,\ \Upsilon_1 \succeq O,\ W_t \succeq O && \text{(A.27a)} \\
\Upsilon_2\,(I - W_t) = O,\ \Upsilon_2 \succeq O,\ (I - W_t) \succeq O && \text{(A.27b)}
\end{aligned}
$$

and

$$
\mu\,\left(\det\left(W_t\right) - \phi_m\right) = 0,\ \mu \geq 0, \det\left(W_t\right) = \phi_m \qquad \text{(A.28)}
$$

We can first rearrange (A.26b)-(A.26d). Note that $\Gamma_3 = Q$ (as implied by (A.26d)) and using the notation $\tilde{K}_t \equiv \Sigma^m_{t|t}\, c\, \left(\Omega^m_{t|t}\right)^{-1} c'$, we can express (A.26b) as

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

and (A.26c) as

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

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 (A.26a) by $W_t\ (I - W_t)$ on the left yields

$$- \left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \Gamma_1 \left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} W_t\ (I - W_t) + \mu\, \phi_m\ (I - W_t) = O \tag{A.29}$$

after applying the slackness conditions (from which $(\Upsilon_1 - \Upsilon_2)\, W_t\ (I - W_t) = O$). Furthermore, we observe the following eigen-decomposition is feasible:

$$\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} \Gamma_1 \left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} = U\, G\, U'$$

That is, it should share the basis with Υ_1, Υ_2 and W_t. Then, the above expression can be written as

$$U\ \left(\mu\, \phi_m\, I - G\ (I - D)\right) D\, U' = O \tag{A.30}$$

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

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. Thus, the eigenvalues stored in G are described as $g_1 \geq g_2 \geq \cdots \geq g_n$ (where n is the dimension of x_t). For $k = 1, \cdots, n$, 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 $k = n$ and it must be $g_n \geq \theta_n$). Using this notation, the ith element of D, d_i, is going to be

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

Thus, the integer k describes the number of eigenvectors that receive positive weights, while the remaining $n - k$ receive a zero weight. 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

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

From this, the optimal $\Sigma^m_{t|t}$ is derived as

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

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 cognitive states. 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 $g_k \geq \theta_k \geq g_{k+1}$.

A.4 A Simpler Approximation

Let's suppose that the memory variable $\tilde{m}_{i,t}$ encodes a linear combination of $x_{i,t|t-1}$. Thus, $\tilde{m}_{i,t}$ is a scalar random variable of the following form.

$$\tilde{m}_{i,t} = \lambda_t\,d_t'\,x_{i,t|t-1} + \tilde{\omega}_{i,t}, \quad \tilde{\omega}_{i,t} \sim \mathcal{N}\left(0,\,\sigma^2_\omega\right) \tag{A.31}$$

Applying the normalization condition $Cov\left[\tilde{m}_{i,t}, d_t'\,x_{i,t|t-1}\right] = Var\left[\tilde{m}_{i,t}\right]$ again pins down the variance of memory noise as

$$\sigma^2_\omega = \lambda_t(1-\lambda_t)\,d_t'\,Var\left[x_{i,t|t-1}\right]d_t$$

The scalar memory variable in (A.31) then results in the prior belief of the form (A.13) where the loading matrix $\tilde{\Lambda}_t$ is described as

$$\tilde{\Lambda}_t = \frac{\lambda_t}{d_t' \, Var\left[x_{i,t|t-1}\right] d_t} Var\left[x_{i,t|t-1}\right] d_t \, d_t', \tag{A.32}$$

and the variance of memory noise is characterized as $\Sigma_\omega = (I - \tilde{\Lambda}_t) \, Var\left[x_{i,t|t-1}\right] \tilde{\Lambda}_t'$ as before. Finally, the mutual information between $\tilde{m}_{i,t}$ and $x_{i,t|t-1}$ is derived as

$$\mathcal{I}\left(\tilde{m}_{i,t}; x_{i,t|t-1}\right) = \mathcal{I}\left(\tilde{m}_{i,t}; d_t' \, x_{i,t|t-1}\right) = h(\tilde{m}_{i,t}) - h(\tilde{m}_{i,t}|d_t' \, x_{i,t|t-1}) = -\frac{1}{2}\log(1-\lambda_t)$$

Applying the memory constraint (1.10) then pins down λ_t as

$$\lambda_t = 1 - \phi_m.$$

B Special Case: A Single State Variable

This section discusses the optimal cognitive process when the only state variable that matters for forecasting is z_t. This is a special case of the setting discussed in the previous section, where the state vector is expressed as x_t and the news vector provides noisy information about $c' x_t$. This section assumes instead that the news vector is expressed as

$$N_t = z_t + \bar{\nu}_t, \quad \bar{\nu}_t \sim \mathcal{N}\left(0, \, \sigma_\nu^2\right). \tag{B.33}$$

B.1 Optimal representation of noisy news

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

$$\tilde{n}_{i,t} = \kappa_{zt} \cdot (z_t + \bar{\nu}_t) + \tilde{u}_{i,t}$$

for some positive scalar κ_{zt} that remains to be specified, and the idiosyncratic noise $\tilde{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(0, \, \sigma_{ut}^2\right)$. The values of κ_{zt} and σ_{ut}^2 that satisfy the normalization assumption (A.2b) and the information constraint (1.7) are derived as below.

$$\kappa_{zt} = \frac{\Sigma_{z,t|t}^m}{\Sigma_{z,t|t}^m + \sigma_\nu^2 + \frac{\phi_n}{1-\phi_n}\left(\sigma_z + \sigma_\nu^2\right)} \tag{B.34}$$

$$\sigma_{ut}^2 = \kappa_{zt}^2 \left(\frac{\phi_n}{1-\phi_n}\right)\left(\sigma_z + \sigma_\nu^2\right) \tag{B.35}$$

where $\sigma_z \equiv Var\left[z_t\right]$.

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 that remains to be specified. The idiosyncratic noise $\tilde{\omega}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(0, \, \sigma_{\omega t}^2\right)$. The normalization assumption (A.2b) and the information constraint (1.10) pin down the values of λ_t and $\sigma^2_{\omega t}$ as below.

$$\lambda_t = 1 - \phi_m \tag{B.36}$$

$$\sigma^2_{\omega t} = \phi_m \left(1 - \phi_m\right) Var\left[z_{i,t|t-1}\right] \tag{B.37}$$

The simplicity of the solution arises because it is not necessary to allocate memory resources to more than one variable.

C Derivations of β_I and β_C

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

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

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

$$z_{t|t} = (1-\lambda)\left(1-\kappa_z\right)\mu + \lambda\,\left(1-\kappa_z\right) z_{t|t-1} + \kappa_z\, z_t + \kappa_z\,\tilde{\nu}_t \tag{C.38}$$

I define b as the weight on unconditional prior belief.

$$b \equiv (1-\lambda)\left(1-\kappa_z\right) \tag{C.39}$$

C.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

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

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

$$\beta_I = -\frac{b}{2\,b + \rho^{-2} - 1} \tag{C.40}$$

Derivation of β_C

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

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

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

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

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

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

Therefore, it is straightforward to see that

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

We can also show that

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

And finally,

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

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

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

Then,

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

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

$$\beta_C = \frac{1-\kappa_z}{\kappa_z}\left(1+(1-\lambda)\,\frac{\lambda(1-\kappa_z)\rho^2}{1-\lambda(1-\kappa_z)\rho^2}\right) \tag{C.41}$$

C.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.

$$\Sigma_{-1} = \rho^2\,\Sigma + \sigma_\epsilon^2 \tag{C.42a}$$
$$\Sigma^m = (1-\lambda)\,\sigma_z^2 + \lambda\,\Sigma_{-1} \tag{C.42b}$$
$$(\Sigma)^{-1} = (\Sigma^m)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} \tag{C.42c}$$

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_z and b are

$$\kappa_z = \frac{\Sigma^m}{\Sigma^m+\tilde{\sigma}_u^2} \tag{C.43}$$

$$b = (1-\lambda)\ \frac{\tilde{\sigma}_u^2}{\Sigma^m+\tilde{\sigma}_u^2} \tag{C.44}$$

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

C.3 Comparative Statics

Comparative Statics for the Uncertainty

Equations (C.42) implicitly impose the following relation.

$$\begin{aligned}
F\left(\Sigma;\tilde{\sigma}_u^2,\lambda\right) &= (\Sigma^m)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - (\Sigma)^{-1} \\
&= \left((1-\lambda)\,\sigma_z^2 + \lambda\,\left(\rho^2\,\Sigma+\sigma_\epsilon^2\right)\right)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - (\Sigma)^{-1} = 0
\end{aligned} \tag{C.45}$$

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

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

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

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

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_z and b

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

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

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_z with respect to $\tilde{\sigma}_u^2$ is:

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

Finally, the derivative of κ_z with respect to λ:

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

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 (C.40) 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

$$\frac{\partial \beta_I}{\partial \phi_m} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \phi_m} = -\frac{\partial \beta_I}{\partial b}\frac{\partial \beta_I}{\partial \lambda} < 0 \tag{C.46a}$$

$$\frac{\partial \beta_I}{\partial \phi_n} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} < 0 \tag{C.46b}$$

Comparative Statics for β_C

Next, we analyze the comparative statics for β_C. The derivative of β_C with respect to ϕ_n is straightforward. From (C.41), we can see that β_C decreases in κ_z, and from above we also know that κ_z decreases in $\tilde{\sigma}_u^2$. Therefore, we have

$$\frac{\partial \beta_C}{\partial \phi_n} = \frac{\partial \beta_C}{\partial \kappa_z}\frac{\partial \kappa_z}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} > 0 \tag{C.47}$$

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

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

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

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

I would like to show that we can find $\hat{\sigma}_u^2$ such that for any λ, the term inside 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 $\kappa_z \to 1$) and negative for $\tilde{\sigma}_u^2 \to \infty$ (since $\Sigma \to \sigma_z^2$ and $\kappa_z \to 0$) 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 $\frac{1-\lambda(1-\kappa_z)\rho^2}{1-\lambda(1-\kappa_z)^2\rho^2}$ decreases in $1-\kappa_z$ (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 $\tilde{\sigma}_u^2 \le \hat{\sigma}_u^2$. In practice, we could find such $\hat{\sigma}_u^2$ by finding $\tilde{\sigma}_u^2$ under which

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

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

D Monetary Model

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

D.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

$$w_t - p_t = \sigma\,c_t + \varphi\,n_t \tag{D.48}$$

$$c_t = E_t\,c_{t+1} - \frac{1}{\sigma}\left(-q_t - E_t\,\pi_{t+1} + log\,\beta\right) \tag{D.49}$$

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

D.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

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

D.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,

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

D.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

$$mc_{t+h} = \chi\, x_t$$

where I define $\chi \equiv \sigma + \varphi$.

D.5 Firms' Macroeconomic Expectations

Substituting (4.7), 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 (4.8) and (4.9), we get

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

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

$$\pi_t = (1-\alpha)\left\{\hat{\kappa_z}\, z_t + \hat{b}\,\hat{\mu}_{t-1}\right\} \tag{D.50}$$

D.6 Inflation Determination

We can solve for the equilibrium inflation process using a guess-and-verify approach. The equation (4.6) states that z_t is determined by π_t and e_t, and the equation (D.50) 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.

$$\pi_t = \varphi_e\, e_t + \varphi_\mu\, \hat{\mu}_{t-1} \tag{D.51}$$

Combining (4.6), (D.50), and (D.51), we can find the coefficients φ_e and φ_μ that verify our initial guess. They are derived as below.

$$\varphi_e = \frac{\delta}{1 + \delta\, \chi\, \frac{\theta}{1-\theta} + \frac{1}{\alpha}\, \frac{1-\hat{\kappa}}{\hat{\kappa}}}$$
$$\varphi_\mu = \frac{1}{1 + \delta\, \chi\, \frac{\theta}{1-\theta} + \frac{1}{\alpha}\, \frac{1-\hat{\kappa}}{\hat{\kappa}}}\; \frac{1-\alpha}{\alpha}\, \frac{\hat{b}}{\hat{\kappa}}$$

where $\delta \equiv \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}$, $\hat{\kappa_z} = \kappa_z + \kappa_\mu$, and $\hat{b} = \lambda(1-\kappa_z) + \frac{\alpha\beta}{1-\alpha\beta}\lambda(1-\kappa_\mu)$.

D.7 Variability of Inflation

From (D.51), 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 (4.6), 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

$$\varpi_e = \left(1 - (1-\alpha\beta)\chi s\right) \varphi_e + (1-\alpha\beta)$$
$$\varpi_\mu = \left(1 - (1-\alpha\beta)\chi s\right) \varphi_\mu$$

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

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

E A Stationary Relationship

The law of motion of the posterior mean for x_t is derived as follows by combining equations (A.13) and (A.17).

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

where $\bar{\nu}_t$ captures the common noise in external information sources, and $noise_{i,t}$ is the idiosyncratic cognitive noise defined as

$$noise_{i,t} \equiv \left(I - \tilde{K}_t\right)\bar{\omega}_{i,t} + \tilde{u}_{i,t}. \tag{E.52}$$

The variance of cognitive-noise term is then derived as

$$Var\left[noise_{i,t}\right] = \left(I - \tilde{K}_t\right)\Sigma_{\omega t}\left(I - \tilde{K}_t\right)' + \Sigma_{ut}.$$

To further ease the notation burden, I introduce the following matrices

$$\Delta_t \equiv \left(I - \tilde{K}_t\right)\left(I - \tilde{\Lambda}_t\right)$$
$$\hat{A}_t \equiv \left(I - \tilde{K}_t - \Delta_t\right)A$$

Using these matrices, the posterior mean $x_{i,t|t}$ has the following law of motion

$$x_{i,t|t} = \Delta_t\, \mu_x + \hat{A}_t\, x_{i,t-1|t-1} + \tilde{K}_t\, x_t + \tilde{\nu}_t + noise_{i,t}. \tag{E.53}$$

The posterior beliefs converge to a stationary distribution as $t \to \infty$. Thus, the matrices such as $\tilde{K}_t$ and $\tilde{\Lambda}_t$, which are functions of the underlying prior and posterior variances, also converge to constant values, whose limits are denoted as $\tilde{K}_t \to \tilde{K}$ and $\tilde{\Lambda}_t \to \tilde{\Lambda}$. After a long enough learning period, the posterior mean for x_t then evolves according to

$$x_{i,t|t} = \Delta\, \mu_x + \hat{A}\, x_{i,t-1|t-1} + \tilde{K} x_t + \tilde{\nu}_t + noise_{i,t} \tag{E.54}$$

The consensus forecasts $x_{t|t}$ average the individual forecasts $x_{i,t|t-1}$ across the continuum of forecaster i in (E.54) at each time t. Thus, $x_{t|t}$ evolves according to

$$x_{t|t} = \Delta\, \mu_x + \hat{A}\, x_{t-1|t-1} + \tilde{K} x_t + \tilde{\nu}_t, \tag{E.55}$$

The deviation of individual forecasts from the consensus can be derived from subtracting (E.55) from (E.54). We can see that it has the following law of motion.

$$x_{i,t|t} - x_{t|t} = \hat{A}\left(x_{i,t-1|t-1} - x_{t-1|t-1}\right) + noise_{i,t}.$$

Thus, the gap is persistent and is affected by a new draw of cognitive noise. Iterating the above equation backward yields that

$$x_{i,t|t} - x_{t|t} = \sum_{j=0}^{\infty} \hat{A}^j\, noise_{i,t-j} \equiv NoiseHistory_{i,t} \tag{E.56}$$

That is, $x_{i,t|t} - x_{t|t}$ captures the accumulated cognitive noise in the past. I compute the stationary variation of $NoiseHistory_{i,t}$ from the following Riccati equation

$$Var\left[NoiseHistory_{i,t}\right] = \hat{A}\, Var\left[NoiseHistory_{i,t}\right]\, \hat{A}' + Var\left[noise_{i,t}\right],$$

which can be computed from the stationary level of $Var\left[noise_{i,t}\right]$.

The law of iterated expectations does not hold. Rearranging (E.54) yields the following expression

$$x_{i,t|t} = x_{i,t|t-1} + \tilde{K}\left(x_t - x_{i,t|t-1}\right) - \Delta\left(x_{i,t|t-1} - \mu_x\right) + \tilde{\nu}_t + noise_{i,t}, \tag{E.57}$$

from which one could see that

$$E\left[x_{i,t|t}\middle|\, \tilde{m}_{i,t-1}, \tilde{n}_{i,t-1}\right] = (I - \Delta)\, x_{i,t|t-1} + \Delta\, \mu_x. \tag{E.58}$$

That is, the law of iterated expectations does not hold as long as Δ is not zero. From the definition of this matrix, we can conclude that if the matrix $\tilde{\Lambda}_t$ is not an identity matrix, Δ will not be a zero matrix. Thus, memory frictions prevent the information set from being nested.

E.1 Perceived Variations of Forecasts

We derive the *perceived* variations of the posterior means that are consistent with DM's prior belief about $x_t \sim \mathcal{N}(\mu_x,\, \Sigma_x)$. Given the stationary values of $\Sigma_{t|t}$, the variations of $x_{i,t|t}$, its covariance with x_t, and its auto-covariance should satisfy the following identities.

$$\begin{aligned}
Var\left[x_{i,t|t}\right] &= \Sigma_x - \Sigma_{t|t} && \text{(E.59a)}\\
Cov\left[x_{i,t|t}, x_t\right] &= Var\left[x_{i,t|t}\right] && \text{(E.59b)}\\
Cov\left[x_{i,t|t}, x_{i,t-1|t-1}\right] &= Cov\left[\hat{A}x_{i,t-1|t-1} + \tilde{K}A\, x_t, x_{i,t-1|t-1}\right]\\
&= Cov\left[\left(\hat{A} + \tilde{K}A\right) x_{i,t-1|t-1}, x_{i,t-1|t-1}\right]\\
&= (I - \Delta)\, A\, Var\left[x_{i,t|t}\right] && \text{(E.59c)}
\end{aligned}$$

where the equation (E.59a) is derived from decomposing the perceived variability of x_t into the (average) variability explained by a given realized cognitive state and the variability of the posterior mean arising due to the randomness in the cognitive states (i.e., the "Law of Total Variance"). The equation (E.59b) uses the "Law of Total Covariance" by focusing on the role of DM's time-t cognitive state. Finally, the auto-covariance (E.59c) uses the law of motion of the posterior mean (E.54).

The same set of statistical properties of the consensus forecast is derived below. (E.56).

$$\begin{aligned}
Var\left[x_{t|t}\right] &= Var\left[x_{i,t|t}\right] - Var\left[NoiseHistory_{i,t}\right] && \text{(E.60a)}\\
Cov\left[x_{t|t}, x_t\right] &= Cov\left[x_{i,t|t}, x_t\right] && \text{(E.60b)}\\
Cov\left[x_{t|t}, x_{t-1|t-1}\right] &= \hat{A}\, Var\left[x_{t|t}\right] + \tilde{K}A\, Cov\left[x_t, x_{t|t}\right] && \text{(E.60c)}
\end{aligned}$$

The derivations are based on its law of motion (E.55) and its relationship with the individual forecasts.

Finally, the variations related to the gap in views $x_{i,t|t} - x_{t|t}$ are derived as below.

$$Var\left[x_{i,t|t} - x_{t|t}\right] = Var\left[NoiseHistory_{i,t}\right] \tag{E.61a}$$

$$Cov\left[x_{i,t|t} - x_{t|t}, x_{i,t-1|t-1} - x_{t-1|t-1}\right] = \hat{A}\,Var\left[x_{i,t|t} - x_{t|t}\right] \tag{E.61b}$$

Forecast errors and revisions. I furthermore discuss the statistical properties of forecast errors and revisions. To reduce the notation burden, I first define the following terms.

$$error_{i,t} = x_t - x_{i,t|t}$$
$$revision_{i,t} = x_{i,t|t} - x_{i,t|t-1}$$

Again, I denote the average errors and forecasts as $error_t$ and $revision_t$, which are averages of $error_{i,t}$ and $revision_{i,t}$ (across the continuum of forecaster i) at each forecasting period t.

At the individual-level, the covariance between forecast errors and revisions is derived as below.

$$Cov\left[error_{i,t}, revision_{i,t}\right] = -Cov\left[error_{i,t}, x_{i,t|t-1}\right] \tag{E.62}$$
$$= -\Delta\,Var\left[x_{i,t|t-1}\right] \tag{E.63}$$

where the first equality holds because an efficient use of time-t information requires the covariance between $error_{i,t}$ and $x_{i,t|t}$ be a zero matrix. The second equality uses the relationship (E.58). Using the definition of $revision_{i,t}$, the variation of forecast revisions is derived as

$$Var\left[revision_{i,t}\right] = Var\left[x_{i,t|t}\right] + Var\left[x_{i,t|t-1}\right] - Cov\left[x_{i,t|t}, x_{i,t|t-1}\right] - Cov\left[x_{i,t|t-1}, x_{i,t|t}\right].$$

All the terms are already derived in the equations (E.59).

To compute the above relationship of the consensus forecast, it is useful to rearrange terms in the law of motion (E.55 the following way.

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

Thus,

$$\tilde{K}\,Cov\left[error_t, revision_t\right]$$
$$= \left(I - \tilde{K}\right)Var\left[revision_t\right] + \Delta\,Cov\left[x_{t|t-1}, revision_t\right] - Cov\left[\tilde{\nu}_t, x_{t|t}\right] \tag{E.64}$$

In our exercise, $\tilde{K}$ is invertible. It remains to derive the variation of $revision_t$ as below.

$$Var\left[x_{t|t} - x_{t|t-1}\right] = Var\left[x_{t|t}\right] + Var\left[x_{t|t-1}\right] - Cov\left[x_{t|t}, x_{t|t-1}\right] - Cov\left[x_{t|t-1}, x_{t|t}\right]$$

whose terms again are computed from equations (E.60).

Using column vectors α_h and α_{h+1} to describe the relationship $y_{t+h} = \alpha_h'\,x_t = \alpha_{h+1}'\,x_{t-1}$, I describe the regression coefficients of interests as below.

$$\beta_C = \frac{Cov\left[y_{t+h} - F_t y_{t+h}, F_t y_{t+h} - F_{t-1} y_{t+h}\right]}{F_t y_{t+h} - F_{t-1} y_{t+h}} \tag{E.65}$$

$$= \frac{\alpha_h' Cov\left[error_t, revision_t\right] \alpha_h}{\alpha_h' Var\left[revision_t\right] \alpha_h} \tag{E.66}$$

$$\beta_I = \frac{Cov\left[y_{t+h} - F_{i,t} y_{t+h}, F_{i,t} y_{t+h} - F_{i,t-1} y_{t+h}\right]}{F_{i,t} y_{t+h} - F_{i,t-1} y_{t+h}} \tag{E.67}$$

$$= \frac{\alpha_h' Cov\left[error_{it}, revision_{it}\right] \alpha_h}{\alpha_h' Var\left[revision_{it}\right] \alpha_h} \tag{E.68}$$

$$\beta_P = \frac{Cov\left[F_{i,t} y_{t+h} - F_t y_{t+h}, F_{i,t-1} y_{t+h} - F_{t-1} y_{t+h}\right]}{F_{i,t-1} y_{t+h} - F_{t-1} y_{t+h}} \tag{E.69}$$

$$= \frac{\alpha_h' Cov\left[x_{i,t|t} - x_{t|t}, x_{i,t-1|t-1} - x_{t-1|t-1}\right] \alpha_{h+1}}{\alpha_{h+1}' Var\left[x_{i,t-1|t-1} - x_{t-1|t-1}\right] \alpha_{h+1}} \tag{E.70}$$

E.2 Parameter Uncertainty: Perceived Variations vs. Actual Variability

In the main text, section 3 discusses the possible uncertainty about the parameter value μ. In this exercise, DM has a correct understanding that the very long-run mean μ is constant but in uncertain about the exact value. Since the true data-generating process stems from a fixed μ, the perceived variations derived so far will be different from the actual variability. To adjust for this different, it is useful to characterize how forecasts are related to this parameter.

For a given value of μ, the expectations of (E.54) are derived as below.

$$E\left[\left. x_{i,t|t} \right| \mu\right] = \Delta\, \mu_x + \hat{A}\, E\left[\left. x_{i,t-1|t-1} \right| \mu\right] + \tilde{K} E\left[\left. x_t \right| \mu\right]$$

Thus, the stationary distribution of the forecast depends on the parameter μ in the following way.

$$E\left[\left. x_{i,t|t} \right| \mu\right] = cons + \underbrace{\left(I - \hat{A}\right)^{-1} \tilde{K}}_{\equiv D} E\left[\left. x_t \right| \mu\right] \tag{E.71}$$

where a matrix D is introduced to capture the loading of $E\left[\left. x_{i,t|t} \right| \mu\right]$ on $E\left[\left. x_t \right| \mu\right]$. Likewise, we can derive the relationship of $error_{i,t}$ and $revision_{i,t}$ to $E\left[\left. x_t \right| \mu\right]$

$$E\left[\left. error_{i,t} \right| \mu\right] = \left(I - D\right) E\left[\left. x_t \right| \mu\right] \tag{E.72}$$

$$E\left[\left. revision_{i,t} \right| \mu\right] = \left(I - A\right) D\, E\left[\left. x_t \right| \mu\right] \tag{E.73}$$

Since the statistics of relevance is the variability given that a parameter μ is fixed, the covariance between forecast errors and revisions is derived as

$$\begin{aligned} Cov\left[\left. error_{i,t}, revision_{i,t} \right| \mu\right] = {} & Cov\left[error_{i,t}, revision_{i,t}\right] \\ & - Cov\left[E\left[\left. error_{i,t} \right| \mu\right], E\left[\left. revision_{i,t} \right| \mu\right]\right], \end{aligned}$$

from the law of total covariance. Using the correction equations (E.72) and (E.73), we can derive the covariance as

$$Cov\left[error_{i,t}, revision_{i,t} \middle| \mu\right] = Cov\left[error_{i,t}, revision_{i,t}\right] \tag{E.74}$$
$$- \left(I - D\right) Var\left[E\left[x_t \middle| \mu\right]\right] D' \left(I - A\right)' \tag{E.75}$$

Likewise, the actual variation in forecast revisions is derived as

$$Var\left[revision_{i,t} \middle| \mu\right] = Var\left[revision_{i,t}\right] - Var\left[E\left[revision_{i,t} \middle| \mu\right]\right]$$
$$= Var\left[revision_{i,t}\right] - \left(I - A\right) D\, Var\left[E\left[x_t \middle| \mu\right]\right] D' \left(I - A\right)' \tag{E.76}$$

The adjustments in (E.75) and (E.76) are made because DM treats the parameter as a random variable (when in fact the data is generated from a fixed parameter). When deriving the corresponding statistics of the average forecasts, we subtract the same terms.

F Estimation

F.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 Both Gross Domestic Product measures are seasonally adjusted, annual rate. Before 1992, forecasts for Nominal GDP (NGDP) correspond to those for nominal Gross National Product. Real GDP (RGDP) is chain-weighted; before 1992, it was fixed-weighted. Before 1981:Q3, forecasts for RGDP are imputed using forecasts for NGDP and GDP Chain-Weighted Price Index (PGDP) as NGDP/PGDP*100.

Real-time Macroeconomic Data

I use the real-time data set provided by the Philadelphia Fed. The third 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, as illustrated in Section 5 of the main text.

Using the real-time data allows us to compare the forecast data to a correct macroeconomic variable with a consistent definition and to compute the errors in forecasts more accurately. Comparing the survey forecasts to the latest release of the data can be misleading. This is because macroeconomic variables are redefined or reclassified, and the base year changes for the real variables. Because 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.