FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Macroeconomic Expectations and Cognitive Noise

Yeji Sung
Federal Reserve Bank of San Francisco

May 2025

Working Paper 2024-19

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

Suggested citation:

Sung, Yeji. 2025. "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

May 6, 2025

Abstract

Standard models of information frictions explain sluggish forecasts but struggle to account for cases of excessive sensitivity. I argue that this limitation arises because these models define information frictions too narrowly—emphasizing constraints on processing new information while assuming past information remains fully accessible. I propose a broader framework that integrates both attention and memory frictions, which reflects constraints on processing new and past information, respectively. This unified model better matches observed forecast biases, enables joint identification of the two frictions from survey data, and provides novel insights for optimal monetary policy responses to cost-push shocks.

JEL codes: D84, E32, E71, G41
Keywords: Information Frictions, Memory Frictions, Expectation Formation, Optimal Monetary Policy

*I am deeply grateful to Michael Woodford, Hassan Afrouzi, Jennifer La'O, and Mark Dean for their invaluable support. I also thank Yueran Ma, Oliver Pfäuti, and Ted Sanders for helpful discussions. In addition, I appreciate feedback from members of the Cognition and Decision Lab, as well as seminar participants at Columbia University, the Federal Reserve Board, the Federal Reserve Bank of St. Louis, the Federal Reserve Bank of San Francisco, and various other seminars and conferences. The views expressed herein are those of the author and do not necessarily represent the views of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

†Email:yeji.sung@sf.frb.org.

Introduction

Economic forecasters face a challenge: how to distill vast amounts of data into coherent predictions. Models of information frictions formalize this challenge by explicitly modeling the constraints on acquiring and processing information (Mankiw & Reis, 2002; Sims, 2003; Woodford, 2003). Because these constraints limit the use of all available information, standard models typically imply that each forecaster relies on a limited and somewhat idiosyncratic information set. Empirical evidence broadly supports these models, confirming their predictions that forecasts adjust gradually rather than instantly to new economic news (Coibion & Gorodnichenko, 2012, 2015) and that disagreement among forecasters persists over time (Goldstein, 2023).

Yet, standard models of information frictions struggle to explain why forecasts sometimes respond excessively—rather than sluggishly—to news. For example, professional forecasts of macroeconomic and financial variables often exhibit excess sensitivity at the individual level, despite consensus forecasts adjusting more slowly (Bordalo et al., 2020). In addition, *long-term* forecasts exhibit heightened sensitivity in diverse settings, including macroeconomic forecasts (Halperin & Mazlish, 2025), firm earnings forecasts (Bordalo et al., 2023), and predictions about longer-maturity interest rates (d'Arienzo, 2020; Wang, 2021). These findings highlight a limitation of standard models of information frictions: their emphasis on sluggish adjustments leaves them unable to explain cases of excessive forecast sensitivity.

I argue that these empirical shortcomings arise because standard models define information frictions too narrowly. Existing frameworks primarily consider frictions in using new information, implicitly assuming that previously acquired information remains fully accessible. Yet, if interpreting new data is cognitively costly and error-prone, why would handling past information be any different? Indeed, recent studies highlight that frictions in retaining and retrieving past information can amplify forecasts' sensitivity to new information (Afrouzi et al., 2023; Azeredo da Silveira & Woodford, 2019; Azeredo da Silveira et al., 2024; Nagel & Xu, 2022). Building on these insights, I propose a unified framework that incorporates cognitive constraints in processing *both* new and past information. This broader approach better matches observed patterns in survey forecasts, enables joint identification of these frictions, and provides new insights for monetary policy.

My model draws on the concept of *cognitive noise*—errors that arise from constraints on information processing (Sims, 2003). Because such noise makes information less reliable, forecasters optimally discount it when forming expectations. These expectations are constrained-optimal but systematically depart from full-information benchmarks. In my model, cognitive noise results from two distinct frictions: *attention frictions,* which limit the processing of new information, and *memory frictions*, which limit the processing of past information. Past information refers to forecasters' prior beliefs and knowledge accumulated from earlier forecasting experiences.

Attention and memory frictions are not meant to suggest that forecasters are simply inattentive or forgetful in a colloquial sense. Rather, they reflect a deeper cognitive bottleneck: interpreting, retaining, and retrieving large volumes of complex information is difficult. Indeed, forecast biases such as sluggish adjustments and excessive sensitivity are consistently observed among professional forecasters, in household surveys (Coibion & Gorodnichenko, 2015), and in controlled laboratory experiments (Afrouzi et al., 2023). The pervasiveness of these forecast biases suggests a common underlying constraint. Cognitive limitations offer a natural explanation—an idea I formalize through attention and memory frictions.

I characterize how a cognitive system optimally allocates limited attention and memory resources to support accurate forecasting. The resulting two-parameter model captures how new and past information should be processed when cognitive resources are limited. Rather than encoding all available information, the system selectively retains what is most predictive of future outcomes. Memory, in particular, is not a passive storage system subject to arbitrary decay: I show that optimal memory compresses past information efficiently—much like principal component analysis—to preserve what is most informative for forecasting. This structure need not be consciously chosen; it reflects how cognitive systems evolve to manage complexity efficiently under constraint.

Second, I show that the joint presence of attention and memory frictions helps reconcile two seemingly contradictory empirical patterns, documented across many macroeconomic and financial variables using data from professional forecasters. On the one hand, average forecasts exhibit systematic *under-revision*: when average forecasts are revised upward, actual outcomes turn out even higher; when revised downward, outcomes fall even further (Coibion & Gorodnichenko, 2015). On the other hand, individual forecasts display the opposite pattern of *over-revision*: upward revisions typically lead to overestimates, and downward revisions to underestimates (Bordalo et al., 2020). While puzzling at first glance, this contrast emerges naturally in a setting with both attention and memory frictions.

This divergence between under- and over-revision arises from how forecasters optimally handle *noisy information* under cognitive constraints. Attention frictions make new information noisy, leading forecasters to discount it, resulting in attenuated revisions. Memory frictions, by contrast, make past information noisy, so that prior beliefs serve as a less precise anchor for interpreting new data. At the individual level, forecasters make large and idiosyncratic revisions to news because their expectations are less anchored to prior beliefs, resulting in forecasts that appear overly sensitive relative to realized outcomes. But when individual forecasts are averaged, these idiosyncratic revisions partially cancel out, revealing the attenuated forecast response to news caused by attention frictions.

My framework also clarifies how attention and memory frictions jointly shape the evolution of forecasts. When only attention frictions are present, responses to news are persistently sluggish. But adding memory frictions changes the picture: the diminished influence of past information increases uncertainty about the current state, amplifying sensitivity to incoming data. However, as memory continues to decay, the influence of that news also fades. The result is a distinct *spike-and-fade* dynamic in forecasts—initially sharp responses that gradually dissipate. These temporal patterns cannot be captured by either friction in isolation and depend on which constraint binds more tightly in a given environment. This raises a natural question: under what conditions does each friction dominate?

Third, I show that the relative importance of each friction varies systematically with the forecast horizon. The core insight is that forecast behavior reflects the relative informativeness of new and past information. At short horizons, forecasters often rely more heavily on recent data, making attention frictions the primary constraint and leading to sluggish responses. At longer horizons, forecasts aim to predict persistent, low-frequency components that are difficult to observe in real time. This places greater weight on past knowledge, amplifying the role of memory frictions and producing the distinctive spike-and-fade dynamics—precisely in line with the survey-based evidence discussed earlier.[1] Thus, my framework explains why forecasts can appear either sluggish or overly reactive, depending on the forecasting horizon and the underlying information environment.

This horizon-dependent sensitivity distinguishes my framework from standard rational expectations and attention-only models, which typically predict no increase in forecast responsiveness at longer horizons. However, the exact degree of forecast responsiveness at any given horizon ultimately depends on the relative magnitudes of attention and memory frictions, making it an inherently empirical question.

Fourth, I estimate attention and memory frictions using survey forecasts for nine macroeconomic variables spanning inflation, output, and interest rates. Identification relies on the

[1] Indirect evidence further supports this increased sensitivity. For example, long-term survey forecasts of output growth and real interest rates exhibit higher volatility compared to estimates of underlying low-frequency trends derived from statistical or structural models (Crump et al., 2023). Similarly, asset prices for longer-maturity display elevated volatility relative to short-maturity prices, a form of excess volatility inconsistent with standard term-structure models (Giglio & Kelly, 2018).

two revision patterns discussed earlier: under-revision in consensus forecasts and over-revision in individual forecasts. Together, these patterns uniquely pin down a pair of friction parameters for each variable. Eight of nine variables exhibit both patterns, enabling joint estimation. One variable (the 3-month Treasury Bill) is excluded because it does not exhibit individual-level over-revision.

The estimated friction parameters imply horizon-dependent forecast biases: short-term expectations adjust sluggishly due to attention frictions, while long-term expectations exhibit spike-and-fade dynamics due to memory frictions. This joint estimation also clarifies a key identification issue: attention frictions appear nearly twice as large as in standard models because the influence of memory frictions is otherwise absorbed by the attention parameter. When memory frictions are excluded, their effect on forecast sensitivity is misattributed to lower attention frictions. Taken together, these findings demonstrate that attention and memory frictions are both quantitatively significant and jointly necessary to explain observed forecast behavior.

Finally, to explore the policy implications of attention and memory frictions, I analyze optimal monetary policy responses to cost-push shocks. I embed the estimated expectations model into a standard New Keynesian framework with price-setting firms that form expectations according to the cognitive frictions estimated from survey forecasts of PCE inflation. Taking firms' expectation formation as given, the central bank optimally sets inflation and output-gap paths. Cost-push shocks generate a familiar policy tradeoff between inflation and output gap stabilization, but cognitive frictions reshape this tradeoff by altering the dynamics of expectation formation.

My analysis shows that attention and memory frictions have distinct implications for the optimal policy response. Attention frictions reduce firms' immediate responsiveness to cost-push shocks, thereby limiting their pass-through into inflation. This diminished sensitivity allows the central bank to stabilize inflation without causing large output losses. Memory frictions, however, introduce additional complexity. While they initially amplify long-term expectations, memory frictions also cause these expectations to fade over time as past information is gradually discounted. This creates a tension for monetary policy: should the central bank tighten aggressively to preempt temporary surges in long-term inflation expectations, or should it pursue a more measured response, anticipating that these spikes will subside on their own?

My results show that cognitive frictions strengthen the case for a "look-through" approach, shifting the optimal policy toward greater tolerance of short-run inflation fluctuations. In contrast, policies derived under rational expectations or attention-only models tend to stabilize inflation more aggressively, but at the cost of deeper and more prolonged downturns. These findings underscore the importance of capturing the key cognitive constraints underlying expectation formation. Neglecting these constraints can cause policymakers to misjudge how expectations respond to shocks, leading to ineffective policies and costly economic disruptions.

Literature Review This paper bridges two previously distinct strands of research on cognitive frictions. One focuses on rational inattention and how individuals allocate limited resources to process new information (Gabaix, 2014; Kohlhas & Walther, 2021; Maćkowiak & Wiederholt, 2009; Pfäuti, 2025; Sims, 2003). The other emphasizes how past experiences and imperfect memory shape current beliefs, showing that perceptions of the present are influenced by how information is encoded and recalled over time (Azeredo da Silveira & Woodford, 2019; Azeredo da Silveira et al., 2024; Bordalo et al., 2024; Malmendier & Nagel, 2016; Malmendier & Wachter, 2021; Nagel & Xu, 2022; Salle et al., 2023). I highlight that both frictions stem from a common source—noisy internal processing—and model them as two dimensions of the same cognitive constraint. By treating attention and memory as structurally unified rather than conceptually separate, the framework reduces the need for ad hoc mechanisms and provides a more coherent theory of belief formation under limited information processing.

A related literature adds behavioral features to noisy information models to explain forecast biases—particularly the underrevision in consensus forecasts and overrevision in individual forecasts documented by Coibion and Gorodnichenko (2015) and Bordalo et al. (2020). These extensions often rely on representative heuristics (Bordalo et al., 2020), model misspecification (Angeletos et al., 2021), or strategic motives (Gemmi & Valchev, 2023). In contrast, my model offers a more parsimonious alternative by grounding both patterns in internal cognitive constraints (Woodford, 2020). Rather than assuming distinct deviations from Bayesian updating for each pattern, I show that limited attention and imperfect memory jointly generate these behaviors within a single rational framework. The model also captures how forecast sensitivity varies across horizons—a key empirical feature that prior models have not accounted for. In doing so, the paper contributes to recent efforts to unify the explanation of survey forecast biases (Enke & Graeber, 2023).

Finally, this paper contributes to the literature on monetary policy implications of biased expectations. A crucial insight is that optimal monetary policy prescriptions depend on modeling empirically realistic long-term inflation expectations. Recent work by Carvalho et al. (2023) explicitly connects monetary policy and long-term inflation expectations through short-run inflation surprises. My model similarly links long-term expectations to short-run inflation surprises, but highlights that these expectations are shaped by two cognitive frictions—attention and memory—which yield a distinctive policy implication: optimal monetary policy supports a "look-through" approach to cost-push shocks, tolerating temporary inflation spikes. This contrasts with prior studies advocating aggressive stabilization to avoid the deanchoring of inflation expectations (Gáti, 2023; Orphanides & Williams, 2006). Although my model also generates heightened sensitivity in expectations, memory frictions cause these spikes to fade quickly, suggesting a less aggressive policy response to cost-push shocks may be optimal to prevent excessive output contractions.

1 Modeling Cognitive Noise in Expectations

This section develops a model of expectation formation with cognitive constraints on processing new and past information. Attention and memory frictions are modeled as information-processing limits that generate random errors in expectations. The framework introduces a two-parameter structure that governs the severity of each constraint. This structure provides the basis for analyzing how expectations deviate from the rational benchmark, and sets up the derivation of forecast dynamics under cognitive noise in the next section.

1.1 Cognitive Noise from Noisy Information Processing

Consider a state vector x_t that follows a linear Markov process with Gaussian innovations. In each period, a decision-maker (DM) forecasts current and future values of x_t, minimizing expected quadratic loss:

$$E\left[\sum_{t=0}^{\infty} \beta^t \left(x_{t+h} - F_{i,t}\, x_{t+h}\right)' \, Q \, \left(x_{t+h} - F_{i,t}\, x_{t+h}\right)\right] \tag{1.1}$$

where $F_{i,t}\, x_{t+h}$ denotes DM i's forecast of x_{t+h}.

A core challenge in forming accurate forecasts lies in diagnosing the current state of the economy. Even with extensive data on GDP, employment, and sales, extracting reliable signals about underlying economic conditions is difficult. The problem is not that the DM lacks a model of how x_t evolves—this is assumed to be known—but that she must interpret imperfect and sometimes contradictory information to infer the current state. To isolate this difficulty, I abstract from model uncertainty and focus on frictions in information processing.

I model this friction as **cognitive noise**—random errors that arise during the mental encoding and interpretation of input information due to limited cognitive precision. Following Sims (2003), I model information processing as a *noisy filter* that compresses input into lower-fidelity representations, limiting the accuracy with which information is incorporated into beliefs. This imprecision can be captured as a mutual information constraint that governs how precisely input information in encoded into mental representations.

This framework applies to two types of information input: "**external data**" (e.g., economic statistics or news) and "**internal data**" (e.g., their own knowledge or memory of past information). Both are filtered through the same noisy processing mechanism. The next sections specify how cognitive noise shapes the interpretation of each type of information.

1.2 A Dual-System Framework: Attention and Memory

I propose modeling the cognitive process as comprising two distinct subsystems: the attention system and the memory system, each processing different types of information. Building upon existing research, I integrate two distinct bodies of literature—the rational inattention literature and studies on imperfect memory—to explore the interaction between attention and memory systems.

1.2.1 Attention Filter: Noisy Processing of External Data

Many publicly available data sources offer partial insights into the underlying state, x_t. This information can be consolidated into a comprehensive news vector, N_t. I assume that x_t and N_t follow a joint normal distribution, which simplifies the process of extracting useful information from N_t to predict x_t and provides a mathematically tractable framework for inference.[2]

As previously discussed, the attention system filters the news vector N_t to generate the mental representation, $n_{i,t}$, for each forecaster. This process can be expressed as:

$$n_{i,t} = K_t \cdot N_t + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}\left(0, \Sigma_{ut}\right) \tag{1.2}$$

Here, K_t is a loading matrix that determines how the information in N_t is processed and weighted by the attention system, and $u_{i,t}$ represents attention noise, which is the random distortion or error in how each forecaster processes the news. The noise is uncorrelated with N_t and idiosyncratic to each forecaster. The matrix Σ_{ut} captures the variance and covariance of the noise terms, and it must be positive semi-definite to ensure the model is statistically well-defined. Aside from this requirement, there are no restrictions on the

[2]Since x_t follows a Markov process, where the future state depends only on the present state and not on the past, only current information about x_t is relevant to the DM's forecasting task.

dimensions or structures of K_t and Σ_{ut}. This allows for flexibility in how DM processes and distorts the external data.

I assume that the precision of the mental representation is constrained as follows:[3]

$$I\left(n_{i,t}; N_t\right) \leq -\frac{1}{2}\log\phi_n \tag{1.3}$$

where $\mathcal{I}\left(n_{i,t};\, N_t\right)$ is the Shannon mutual information between $n_{i,t}$ and N_t, which reflects the precision of the attention filter. The parameter $\phi_n \in (0,1)$ constrains the upper bound for this mutual information. As ϕ_n approaches zero, the mutual information tends to infinity, meaning that the mental representation is perfectly precise and contains no noise. Conversely, a larger value of ϕ_n reduces the mutual information, implying a noisier mental representation with less precision.

1.2.2 Memory Filter: Noisy Processing of Internal Data

In addition to processing external data, the DM builds knowledge over time through repeated forecasting experiences. This accumulated knowledge helps the DM refine their predictions about the state of the economy. I denote the internal data available to the DM at time t as $M_{i,t}$, a memory vector composed of two elements:

$$M_{i,t} = \begin{pmatrix} m_{i,t-1} \\ \\ n_{i,t-1} \end{pmatrix} \tag{1.4}$$

Here, $m_{i,t-1}$ represents knowledge carried forward from earlier periods—an internal state that evolves over time and reflects the DM's prior understanding of the economy. The second component, $n_{i,t-1}$, captures the information gained from the previous period's news vector N_{t-1}, providing insights from more recent news. Taken together, these components form the DM's stock of knowledge available at time t.

The memory system operates analogously to the attention system, filtering internal data to form a mental representation of prior knowledge:

$$m_{i,t} = \Lambda_t \cdot M_{i,t} + \omega_{i,t}, \quad \omega_{i,t} \sim \mathcal{N}\left(0,\, \Sigma_{\omega,t}\right) \tag{1.5}$$

[3]The proposed cost function differs from the standard approach in the rational inattention literature. Typically, it is assumed that the DM can arrange to receive a signal, $n_{i,t}$, at time t, conditioned on all the prior signals up to time $t-1$. In that case, the cost is proportional to $\mathcal{I}\left(n_{i,t}; N_t | n_{i,t-1}, \cdots, n_{i,0}\right)$. However, I consider an environment where past realizations of $n_{i,t}$ are not freely available, as the rest of the model will clarify. Thus, I assume that the processing of external data occurs without being influenced by the DM's past mental states. In other words, the mental representation at time t, $n_{i,t}$, is formed based solely on the current information, N_t, and does not depend on or incorporate past realizations of $n_{i,t-1}, n_{i,t-2}, \cdots$.

In this equation, Λ_t is a loading matrix that transforms the internal data in $M_{i,t}$ into a mental representation. The term $\omega_{i,t}$ represents memory noise, capturing random errors in how the past knowledge is recalled and processed. This noise is assumed to be uncorrelated with $M_{i,t}$ and normally distributed, with its variance and covariance described by the matrix $\Sigma_{\omega,t}$. As with the attention filter, $\Sigma_{\omega,t}$ must be positive semi-definite, but no additional restrictions are placed on the dimensions or structure of Λ_t or $\Sigma_{\omega,t}$.

To limit the precision of the memory system, I impose the following constraint on the mutual information between $m_{i,t}$ and $M_{i,t}$:

$$I\left(m_{i,t}; M_{i,t}\right) \leq -\frac{1}{2} \log \phi_m \tag{1.6}$$

where $I\left(m_{i,t}; M_{i,t}\right)$ again represents the Shannon mutual information between $m_{i,t}$ and $M_{i,t}$. The parameter ϕ_m controls the upper bound of this mutual information, effectively setting a limit on how accurately the memory system can capture past knowledge.

When ϕ_m approaches zero, the mutual information increases, implying that the memory system has near-perfect recall of past observations $(n_{i,0}, n_{i,1}, \cdots, n_{i,t-1})$. In this case, the memory vector is fully "nested," meaning the most recent memory vector fully incorporates all previous ones (i.e., $M_{i,t} \supseteq M_{i,t-1}$). On the other hand, as ϕ_m approaches 1, the memory becomes increasingly dominated by noise, resulting in little to no useful information being retained about past states. In this case, the mental representation provides limited predictive value for future states.

Assumptions on Attention and Memory. I model attention and memory systems as operating in parallel, each subject to its own accuracy constraint. This abstraction is motivated by findings in cognitive neuroscience suggesting that attention and memory rely on distinct neural systems.[4] While this separation captures core functional differences, it remains a stylized representation that abstracts away from many complexities of real-world information processing. Still, the model allows for interaction between the two systems: attention shapes what is remembered, and memory in turn guides what we attend to.

1.2.3 Implications of the Linear-Gaussian Filtering Systems

Having described how external and internal data are mentally represented, I now turn to how forecasters integrate these cognitive states. I assume they do so in a manner consistent with Bayesian principles, optimally combining their information to minimize forecast

[4] Attention is largely supported by frontal and parietal regions involved in goal-directed processing, while memory—particularly episodic encoding and retrieval—relies on medial temporal lobe structures such as the hippocampus (Purves et al., 2013).

errors. This yields Bayesian efficiency: all available data—as processed through noisy filters—are used as effectively as possible. For clarity, I refer to $n_{i,t}$ as *noisy news* and $m_{i,t}$ as *noisy memory*.

The linear-Gaussian structure of noisy news ($n_{i,t}$) and noisy memory ($m_{i,t}$) implies that the DM's beliefs about the state vector x_t—both past and present—are Gaussian. Specifically, the conditional distributions $(x_0, \cdots, x_t)\,|m_{i,t}$ and $(x_0, \cdots, x_t)\,|m_{i,t}, n_{i,t}$ are Gaussian, with their second moments reflecting the perceived uncertainty shaped by attention and memory noise.[5] Given the Gaussian nature of beliefs, the DM's forecasts of the future values are also Gaussian. These forecasts are computed using the Kalman filter, which preserves Bayesian efficiency under the linear-Gaussian framework.

To formalize this, I introduce notation for the DM's beliefs about the state x_τ for any time τ—past, present, or future— conditioned on their cognitive states at time t:

$$x_\tau|\, m_{i,t} \sim \mathcal{N}\,(x^m_{i,\tau|t},\, \Sigma^m_{\tau|t}) \tag{1.7}$$

$$x_\tau|\, m_{i,t}, n_{i,t} \sim \mathcal{N}\,(x_{i,\tau|t},\, \Sigma_{\tau|t}) \tag{1.8}$$

The first distribution represents the **prior belief** at the beginning of period t, based solely on the memory state $m_{i,t}$. The superscript m indicates that this belief is memory-based. The second distribution is the **posterior belief** after incorporating the noisy news $n_{i,t}$ and is denoted without the superscript.

The average losses from inaccurate forecasts are thus proportional to $\Sigma_{t|t}$, which reflects the uncertainty in the DM's posterior belief. As a result, the loss function (1.1) simplifies to

$$\sum_{t=0}^{\infty} \beta^t\, trace\,(\Sigma_{t|t}\, \tilde{Q}), \tag{1.9}$$

where $\tilde{Q}$ is a weighting matrix that reflects the cost of forecast errors. It is derived from the law of motion of the state vector and the original weighting matrix Q in equation (1.1). This reduced-form expression makes clear how posterior uncertainty directly drives the expected losses over time.

The sequence $\{K_t, \Sigma_{ut}, \Lambda_t, \Sigma_{\omega t}\}_{t=0}^{\infty}$ fully characterizes the attention and memory processes over time. I assume that these representational systems are **jointly optimized** to minimize the loss function in (1.9), subject to the information constraints in (1.3) and (1.6). This joint optimization captures the interdependence between attention and memory: the availability of external data informs what past knowledge should be stored or

[5]Because of these cognitive frictions, the DM's subjective probability distribution diverges from the true data-generating process.

retrieved, while the accuracy of recalled information shapes how attention is directed in future periods.[6]

1.3 Optimal Information Selection

Given the costs associated with processing information, the DM cannot encode all available data. Instead, the cognitive process must selectively focus on the subset of information that is most relevant to the task at hand. In other words, the attention and memory system extract and retain the most important features, as formalized in the proposition below.

Proposition 1. *Under optimal information selection, the cognitive process follows the structure:*

$$m_{i,t} = \tilde{\Lambda}_t \, E\left[x_t \middle| M_{i,t}\right] + \tilde{\omega}_{i,t}, \quad \tilde{\omega}_{i,t} \sim \mathcal{N}\left(0, \, \tilde{\Sigma}_{\omega t}\right) \tag{1.10}$$

$$n_{i,t} = \tilde{K}_t \, E\left[x_t \middle| N_t\right] + \tilde{u}_{i,t}, \quad \tilde{u}_{i,t} \sim \mathcal{N}\left(0, \, \tilde{\Sigma}_{ut}\right) \tag{1.11}$$

Here, $\tilde{\Lambda}_t$ and $\tilde{K}_t$ are loading matrices, and the terms $\tilde{\Sigma}_{\omega t}$ and $\tilde{\Sigma}_{ut}$ denote the variance-covariance matrices of the cognitive noise components $\tilde{\omega}_{i,t}$ and $\tilde{u}_{i,t}$, respectively. The precise structure of these matrices is derived later, based on the constraints imposed by the information environment.

Proof. The proof is provided in Appendix A.1. □

As outlined in the proposition, both attention and memory retain only the information that improves forecasts of the state vector—specifically, the conditional expectation of x_t given each available information set. Any information that does not affect the conditional expectations merely increases the cognitive cost, measured by mutual information, without improving forecast accuracy. This selective process allows the DM to minimize cognitive load while preserving only the components of information that are directly relevant for prediction.

This structure implies that cognitive states are formed by weighting forecast-relevant expectations through matrix loadings ($\tilde{\Lambda}_t$, $\tilde{K}_t$), while their accuracy determined by the associated noise variances ($\tilde{\Sigma}_{\omega t}$, $\tilde{\Sigma}_{ut}$). However, without further restrictions, different combinations of loadings and noise variances could generate identical beliefs about the state, leading to an indeterminacy. To resolve this, I normalize the system so that each cognitive state ($m_{i,t}$ and $n_{i,t}$) directly serves as the best predictor of x_t (up to a constant).[7]

[6] Although this analysis assumes a linear-Gaussian structure, the underlying principles extend naturally to more general settings.

[7] Specifically, $\mathbb{E}[x_t \mid m_{i,t}] = m_{i,t} + \varepsilon_t$, where ε_t is orthogonal to $m_{i,t}$, and $\mathbb{E}[x_t \mid m_{i,t}, n_{i,t}] = n_{i,t} + \varepsilon_{m,t}$, where $\varepsilon_{m,t}$ depends only on $m_{i,t}$.

Formally, this normalization imposes the following orthogonality conditions:

$$Cov\left[E\left[x_t|\,M_{i,t}\right]-m_{i,t},m_{i,t}\right]=0 \tag{1.12}$$

$$Cov\left[E\left[x_t|\,N_t\right]-n_{i,t},n_{i,t}|\,m_{i,t}\right]=0 \tag{1.13}$$

These constraints ensure that noise variances are uniquely determined by the corresponding matrix loadings: $\tilde{\Sigma}_{\omega,t}$ is fully determined by $\tilde{\Lambda}_t$, and $\tilde{\Sigma}_{u,t}$ by $\tilde{K}_t$.[8] The matrices $\tilde{\Lambda}_t$ and $\tilde{K}_t$ must be chosen so that the resulting noise covariances, $\tilde{\Sigma}_{\omega t}$ and $\tilde{\Sigma}_{ut}$, are valid—that is, symmetric and positive semi-definite.

Even after cognitive constraints in (1.3) and (1.6) are satisfied, the matrix loadings $\tilde{\Lambda}_t$ and $\tilde{K}_t$ can be further refined to improve how information is represented. This may involve applying transformations such as dimensionality reduction or rotations, which help express the selected information more compactly. These transformations must still comply with the original information constraints. Section 3 provides further details on this refinement process.

Now that we have considered how information is selectively encoded, the next step is to determine how this affects the DM's beliefs about the state. To do so, I introduce a simplifying assumption about the information revealed by external data. Specifically, I assume that external data reveals a noisy version of the true state:

$$E\left[x_t|\,N_t\right]=x_t+\nu_t, \quad \nu_t\sim\mathcal{N}\left(0,\,\Sigma_\nu\right) \tag{1.14}$$

This assumption models the conditional mean $E\left[x_t|\,N_t\right]$ as the true state perturbed by normally distributed noise ν_t. Importantly, unlike the cognitive noise discussed earlier, this noise term is embedded in the external data itself and is shared across all forecasters.

Given the structure of the external data in (1.14), the following proposition characterizes how the DM updates beliefs about the state, subject to cognitive constraints.

Proposition 2. *Under optimal information selection, the DM's prior and posterior beliefs at time t are given as follows. The prior belief, conditioned on noisy memory, is:*

$$x^m_{i,t|t}=(I-\tilde{\Lambda}_t)\,\mu_x+\tilde{\Lambda}_t\,x_{i,t|t-1}+\omega_{i,t} \tag{1.15a}$$

$$\Sigma^m_{t|t}=(I-\tilde{\Lambda}_t)\,\Sigma_x+\tilde{\Lambda}_t\,\Sigma_{t|t-1} \tag{1.15b}$$

where $x_t\sim\mathcal{N}\left(\mu_x,\,\Sigma_x\right)$ is the DM's default prior over the state.

[8] Specifically, $\tilde{\Sigma}_{\omega,t}=(I-\tilde{\Lambda}_t)Var\left[E\left[x_t|\,M_{i,t}\right]\right]\tilde{\Lambda}'_t$ and $\tilde{\Sigma}_{u,t}=Cov\left[x_t,E\left[x_t|\,N_t\right]|\,m_{i,t}\right]\tilde{K}'_t-\tilde{K}_t\,Var\left[E\left[x_t|\,N_t\right]|\,m_{i,t}\right]\tilde{K}'_t$.

When noisy news is incorporated, the posterior belief is updated as follows:

$$x_{i,t|t} = (I - \tilde{K}_t)\, x^m_{i,t|t} + \tilde{K}_t\, (x_t + \nu_t) + u_{i,t} \tag{1.16a}$$

$$\Sigma_{t|t} = (I - \tilde{K}_t)\, \Sigma^m_{t|t} \tag{1.16b}$$

Proof. The proof is provided in Appendix A.2. □

The first set of equations characterizes the DM's prior belief, formed using only noisy memory. Importantly, this prior does not simply carry forward the DM's earlier posterior; instead, it partially reverts toward the default prior. The extent of this reversion—visible in both the mean and variance—is governed by the matrix $I - \tilde{\Lambda}_t$. This captures how memory frictions weaken the influence of past knowledge. The smaller $\tilde{\Lambda}_t$, the less precisely past information is carried into current beliefs.

The second set of equations describes the posterior update after incorporating noisy news. Here, the DM adjusts the memory-based prior toward a signal about the current state, with the weight on this new information determined by $\tilde{K}_t$. As with memory, attention frictions shape both the mean and variance of the posterior distribution: the smaller $\tilde{K}_t$, the more muted the response to new information. Together, memory and attention determine how past knowledge and present data jointly shape beliefs about the state.

Scope of the Model. The model characterizes optimal attention and memory systems specifically for the task of forming expectations about a state vector that evolves as a Markov process. Their structure is derived endogenously, given this specific environment and goal. As such, the model abstracts from many other aspects of attention and memory—such as semantic knowledge, cue-driven retrieval, or episodic recall[9]—that matter in other contexts but are less relevant when the sole objective is to forecast temporally linked states. It also does not capture how some experiences become disproportionately memorable—such as emotionally salient events—because in this framework, past observations matter only to the extent that they help predict the current state. The resulting structure should thus be interpreted as task-specific and normative, not as a general-purpose account of cognitive systems.

[9]See, for example, Bordalo et al. (2022), Wachter and Kahana (2023), Wachter and Kahana (2024), Bordalo et al. (2024), Enke et al. (2024) for recent work that incorporates richer mechanisms and models of memory.

2 Cognitive Noise and Forecast Biases

This section derives closed-form expressions that link attention and memory frictions to two key empirical moments: forecast errors regressed on forecast revisions at the consensus and individual levels. These expressions show that both coefficients depend on the joint influence of attention and memory frictions, complicating standard interpretations. The simplified model, with a single-variable state, isolates the core mechanisms and motivates an estimation strategy based on observed forecast bias in survey data.

2.1 Optimal Cognitive Process: Learning a Single State Variable

Consider a univariate latent variable z_t that follows an autoregressive process:

$$z_t = (1-\rho)\,\mu + \rho\, z_{t-1} + \epsilon_t \tag{2.1}$$

where μ is the long-run mean of z_t, and ρ is the serial correlation parameter, with $|\rho| < 1$ ensuring stationarity. The innovation term ϵ_t is independently drawn from a Gaussian distribution, $\mathcal{N}\left(0,\, \sigma_\epsilon^2\right)$. Throughout this section, I assume that the DM has full knowledge of the data-generating process, including the parameters μ, ρ, and σ_ϵ^2.

The state vector in this case consists of a single variable, expressed as $x_t = (z_t)$. This simplification ensures that the optimally selected information is one-dimensional, eliminating the need for dimensionality reduction in the mental representation. Under this setup, the structure of noisy memory and noisy news can be derived in closed form, as outlined in Proposition 1:

$$m_{i,t} = \lambda_t \cdot E\left[z_t |\, M_{i,t}\right] + \omega_{i,t}, \quad \omega_{i,t} \sim \mathcal{N}\left(0,\, \sigma_{\omega t}^2\right) \tag{2.2}$$

$$n_{i,t} = \kappa_t \cdot E\left[z_t |\, N_t\right] + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}\left(0,\, \sigma_{ut}^2\right) \tag{2.3}$$

In the univariate setting, the loading terms λ_t and κ_t are now scalars rather than matrices. The noise variances take closed-form expressions under the normalization conditions from (1.12) and (1.13).[10]

External data continue to provide a noisy signal of the latent state:

$$E\left[z_t |\, N_t\right] = z_t + \tilde{\nu}_t, \quad \tilde{\nu}_t \sim \mathcal{N}\left(0,\, \sigma_\nu^2\right) \tag{2.4}$$

Again, $\tilde{\nu}_t$ captures the intrinsic imprecision of the signal itself—distinct from cognitive noise.

The precision of the mental representation is quantified by its mutual information with

[10] Specifically, $\sigma_{\omega t}^2 = \lambda_t\,(1-\lambda_t)\, Var\left[E\left[z_t |\, M_{i,t}\right]\right]$ and $\sigma_{ut}^2 = \kappa_t\,(1-\kappa_t)\, Var\left[z_t |\, m_{i,t}\right] - \kappa_t^2\,\sigma_\nu^2$.

the input information. For memory, the mutual information between noisy memory $m_{i,t}$ and the memory vector $M_{i,t}$ depends solely on the loading λ_t, and is given by:

$$I\left(m_{i,t}; M_{i,t}\right) = -\frac{1}{2}\log\left(1-\lambda_t\right) \tag{2.5}$$

An analogous expression holds for attention. The mutual information between noisy news $n_{i,t}$ and the news vector N_t is:

$$I\left(n_{i,t}; N_t\right) = -\frac{1}{2}\log\frac{\Sigma^m_{z,t|t} - \kappa_t\left(\Sigma^m_{z,t|t} + \sigma^2_\nu\right)}{\Sigma^m_{z,t|t} + \kappa_t\left(\Sigma_z - \Sigma^m_{z,t|t}\right)} \tag{2.6}$$

where $\Sigma^m_{z,t|t}$ is the DM's prior uncertainty about z_t, and Σ_z is its unconditional prior uncertainty. Since the uncertainty terms are fixed prior to the arrival of new data, this mutual information is also uniquely determined by κ_t.

The loading terms λ_t and κ_t are determined by the cognitive constraints in equations (1.3) and (1.6), which place upper bounds on the mutual information between cognitive states and their corresponding inputs. Because mutual information increases with the loadings, the cognitive constraints impose upper bounds on λ_t and κ_t. Since higher loadings yield more accurate forecasts, the optimal loadings are set at these bounds:

$$\lambda_t = 1 - \phi_m \tag{2.7}$$

$$\kappa_t = \frac{\Sigma^m_{z,t|t}}{\Sigma^m_{z,t|t} + \sigma^2_\nu + \frac{\phi_n}{1-\phi_n}\left(\Sigma_z + \sigma^2_\nu\right)} \tag{2.8}$$

More restrictive cognitive constraints, indicated by higher ϕ_m and ϕ_n, result in lower loadings and less accurate mental representations.

2.2 Channels of Attention and Memory Frictions

Based on the optimal cognitive process, the evolution of beliefs about the latent state z_t is given by:

$$z_{i,t|t} = (1-\lambda)(1-\kappa)\,\mu + \underbrace{\lambda(1-\kappa)}_{\text{Stickiness}} z_{i,t|t-1} + \underbrace{\kappa}_{\text{Sensitivity}} z_t + noise_t + noise_{i,t} \tag{2.9}$$

where λ and κ denote the constant memory and attention loadings, respectively, in a steady state.[11] The two noise terms capture distinct sources of random errors. The term $noise_t \equiv \kappa\,\tilde{\nu}_t$ reflects the common noise arising from imperfect external data. The term $noise_{i,t} \equiv (1-\kappa)\,\omega_{i,t} + u_{i,t}$ captures idiosyncratic cognitive noise, stemming from imprecision in memory and attention filters.

[11] As shown in equation (2.7), λ_t is constant over time, so the time subscript is omitted.

This equation illustrates two key aspects of forecasts: **sensitivity** and **stickiness**. Sensitivity, governed by the Kalman gain κ, measures how strongly beliefs respond to new information. Stickiness, captured by the term $\lambda(1-\kappa)$, reflects the extent to which current forecasts remain anchored to past beliefs.

Attention and memory frictions shape forecast sensitivity and stickiness in distinct ways: while attention frictions limit responsiveness to new data, memory frictions weaken reliance on past knowledge.

Attention frictions operate entirely through the Kalman gain κ, which determines how new information is incorporated into beliefs. High attention frictions cause agents to process external data less precisely, lowering κ and making forecasts more dependent on past beliefs. This mechanically increases stickiness, generating the standard inverse relationship between sensitivity and stickiness observed in noisy information models.

Memory frictions, in contrast, affect forecasts through two distinct channels. First, it directly reduces λ, diminishing the influence of past beliefs and thereby lowering stickiness. Second, memory frictions indirectly increase the Kalman gain by increasing prior uncertainty. Specifically, lower λ raises $\Sigma^m_{z,t|t}$.[12] The resulting higher κ further increases sensitivity and reduces stickiness. These dynamics produce a distinctive **spike-and-decay dynamics** associated with memory frictions: forecasts initially spike but decline over time as the influence of new information fades. This pattern contrasts with dynamics arising purely from attention frictions.

2.3 Biases in Survey Forecasts

This section examines how survey forecasts may deviate from the Full-Information Rational Expectations (FIRE) benchmark. By revisiting standard empirical tests from the literature, I access the extent to which cognitive noise can account for systematic forecast errors.

2.3.1 Tests of Rational Expectations

The FIRE hypothesis relies on three assumptions: (i) **Bayesian efficiency**—all available information is optimally used, assuming correct beliefs about the environment; (ii) **Perfect memory**—past information is accurately retained and utilized; and (iii) **Perfect attention**— new information is accurately utilized.

Regression specifications proposed in the literature implicitly test these assumptions.Bordalo et al. (2020) use individual-level forecasts to examine departures from Bayesian efficiency

[12]Via the equation $\Sigma^m_{z,t|t} = \Sigma_{z,t|t-1} + (1-\lambda_t)(\Sigma_z - \Sigma_{z,t|t-1})$.

and perfect memory:

$$z_{t+h} - F_{i,t}\, z_{t+h} = \alpha_I + \beta_I \left(F_{i,t}\, z_{t+h} - F_{i,t-1}\, z_{t+h}\right) + e_{i,t+h|t}, \tag{2.10}$$

where $F_{i,t}\, z_{t+h}$ denotes forecaster i's expectations at time t. Under these assumptions, forecast revisions should be uncorrelated with subsequent errors ($\beta_I = 0$), as forecasters would already be optimally incorporating their past knowledge.

Similarly, Coibion and Gorodnichenko (2015) use consensus forecasts in a regression test that tests all three FIRE assumptions:

$$z_{t+h} - F_t\, z_{t+h} = \alpha_C + \beta_C \left(F_t\, z_{t+h} - F_{t-1}\, z_{t+h}\right) + e_{t+h|t}, \tag{2.11}$$

where $F_t\, z_{t+h}$ is the average forecast across individuals. If forecasters are Bayesian efficient and process both past and current information without distortion, revisions should not predict future errors ($\beta_C = 0$).

2.3.2 Consensus Error-Revision Test

Coibion and Gorodnichenko (2015) find a positive relationship between consensus forecast errors and revisions ($\beta_C > 0$) across macroeconomic variables, which they attribute to information frictions—such as noisy signals or delayed information—leading to sluggish forecast adjustment. The magnitude of β_C is interpreted to reflect the severity of information frictions.

The cognitive-noise model offers a novel interpretation of β_C by showing how memory frictions contribute to sluggish forecast adjustment.

Proposition 3. *In the presence of attention and memory frictions, the asymptotic limit of β_C is:*

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

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

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

Proof. See Appendix D. □

The proposition reveals two key insights. First, it confirms Coibion and Gorodnichenko (2015)'s core finding: attention frictions generate positive β_C through sluggish updating. Second, it highlights a novel countervailing force—memory frictions. By increasing uncertainty about prior knowledge, memory frictions raise the Kalman gain (κ), leading forecasters to place greater weight on new information and thereby reducing β_C.

2.3.3 Individual Error-Revision Test

Bordalo et al. (2020) document a negative β_I across a range of macroeconomic and financial variables. Under the implicit assumption of perfect memory, standard information frictions imply $\beta_I = 0$ as long as forecasters incorporate their past knowledge efficiently. To explain the negative β_I, the authors thus relax Bayesian efficiency and propose "diagnostic expectations," a model in which agents overweight recent information and revise forecasts excessively.

I offer an alternative interpretation that preserves Bayesian efficiency but relaxes the assumption of perfect memory.

Proposition 4. *In the presence of attention and memory frictions, the asymptotic limit of β_I is:*

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

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

1. *$\beta_I < 0$ if $\phi_m > 0$, and $\beta_I = 0$ if $\phi_m \to 0$.*
2. *$\frac{\partial \beta_I}{\partial \phi_n} < 0$, and $\frac{\partial \beta_I}{\partial \phi_m} < 0$.*

Proof. See Appendix D. □

The proposition demonstrates that memory frictions can generate a negative β_I while preserving Bayesian efficiency. In this framework, the regression coefficient reflects a bias stemming from imperfect use of past knowledge. When forecasters give insufficient weight to accumulated knowledge, they lose their baseline anchors, causing expectations to drift away from relevant prior beliefs. This loss of anchoring leads forecasters to revise disproportionately in the direction indicated by new information.

The magnitude of this bias depends on the interaction between attention and memory frictions. Memory frictions lead forecasters to discount their prior knowledge, pulling β_I downward. This effect intensifies as attention frictions increase: when agents face greater difficulty processing new information, they rely more heavily on their prior knowledge. As a result, their ability to correct memory-driven distortions weakens, and β_I becomes increasingly negative.

2.3.4 Joint Identification of the Cognitive Frictions

The previous propositions show how attention and memory frictions shape forecast biases. Taken together, they provide a basis for identifying the underlying severity of these frictions.

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 D. □

This lemma establishes a one-to-one mapping between the regression coefficients (β_C, β_I) and the underlying cognitive frictions (ϕ_n, ϕ_m). The key to identification lies in the distinct ways these frictions affect each coefficient. To illustrate this, I construct isocurves—sets of (ϕ_n, ϕ_m) pairs that generate a constant value of β_C and β_I.

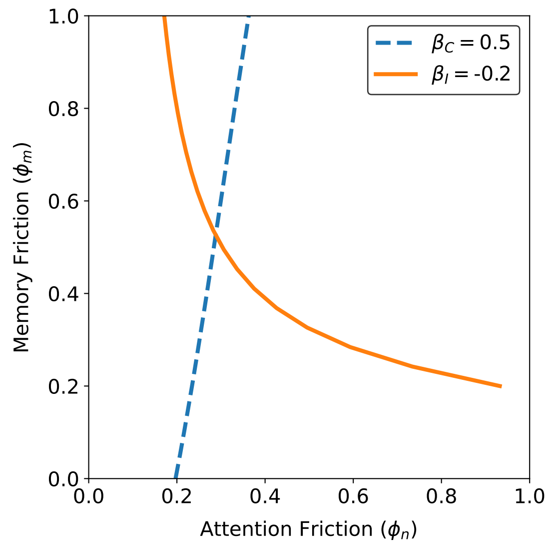
The isocurves exhibit opposing slopes. The β_C isocurve slopes upward: the sluggish updating from higher attention frictions can be offset by increased belief sensitivity from higher memory frictions. In contrast, the β_I isocurve slopes downward: stronger memory frictions, which weaken the influence of past knowledge, can be counterbalanced by more precise processing of external data (i.e., lower attention frictions). These opposing slopes ensure a unique intersection—if one exists—enabling point identification.

Figure 1 illustrates this identification using $\rho = 0.8$ and $\sigma_\epsilon^2 = 1.0$, with σ_ν^2 calibrated to be 2 percent of the variance of z_t.[13] The isocurve for $\beta_C = 0.5$ (solid blue) slopes upward, while the isocurve for $\beta_I = -0.2$ (dashed orange) slopes downward. Their intersection at $\phi_n = 0.29$ and $\phi_m = 0.53$ pinpoints the unique combination of cognitive frictions consistent with both observed biases.

Preview: Horizon-Dependent Biases In the current sections' setting, β_C and β_I are constant across forecast horizons. In the next section, I extend the model to incorporate long-run uncertainty, providing a natural setting to examine how forecast biases evolve across horizons. As the forecast horizon lengthens, memory frictions become more influential, driving both β_C and β_I to become increasingly negative, as detailed in Appendix Section E.

[13]Calibrating σ_ν^2 at 2 percent of the variance of the state is typical for the macroeconomic variables analyzed in Section 4.

Figure 1: Joint identification of cognitive friction using β_C and β_I



The figure displays isocurves for the two regression coefficients in equations (2.11) and (2.10). The solid blue curve shows pairs of attention friction (ϕ_n) and memory friction (ϕ_m) that yield $\beta_C = 0.5$. The dashed orange curve shows pairs that yield $\beta_I = -0.2$. The intersection of the two curves identifies the unique pair of cognitive frictions consistent with both empirical moments.

3 Implications for the Term Structure of Expectations

This section extends the model to distinguish between more and less observable components of the economic state. This richer environment provides a framework for analyzing how the relative influence of each friction varies with the forecast horizon.

3.1 Optimal Cognitive Process: Learning the Term Structure

To account for the uncertainty about the long run, I modify the data-generating process for z_t as follows:

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

$$\mu_t = (1 - \rho_\mu)\, \mu + \rho_\mu\, \mu_{t-1} + \epsilon_{\mu,t}. \tag{3.1b}$$

Here, the exogenous state z_t fluctuates around a time-varying mean μ_t, which evolves stochastically. This contrasts with the previous section, where the mean was fixed. As a result, the DM mut form beliefs about both z_t and μ_t. I therefore define the state variable for prediction as $x_t = (z_t, \mu_t)$

To characterize how the DM learns about the state x_t, I now derive the optimal cognitive process. As characterized by Proposition 1, this requires specifying two sequences of loading matrices: $\tilde{K}_t$ for the attention system and $\tilde{\Lambda}_t$ for the memory system. I use the matrix notation introduced in equations (1.7) and (1.8). Continuing from Section 2.1, I assume that external data provides a noisy signal only about z_t, as specified in equation (2.4).

Optimal Attention Loading Matrices: Deriving $\tilde{K}_t$ When forecasters are uncertain only about the current hidden state z_t, with no long-run uncertainty, Section 2.1 shows that the optimal attention system is one-dimensional. In that case, a single scalar κ_{zt} fully characterizes the attention system, as derived in equation (2.8).

This structure extends naturally to the expanded state space. Because external data informs μ_t only indirectly through z_t, all relevant information about the long-run state contained in N_t is already embedded in $E\left[z_t | N_t\right]$. As a result, the full attention matrices $\tilde{K}_t$ and $\tilde{\Sigma}_{ut}$ in equation (1.11) can be constructed directly from κ_{zt} and σ_{ut}^2, as shown below.

$$\tilde{K}_t = \frac{\kappa_{zt}}{e_1' \, \Sigma_{t|t}^m \, e_2} \, \Sigma_{t|t}^m \, e_2 \, e_2' \tag{3.2}$$

$$\tilde{\Sigma}_{ut} = \frac{\sigma_{ut}^2}{(e_1' \, \Sigma_{t|t}^m \, e_2)^2} (\Sigma_{t|t}^m \, e_2)(\Sigma_{t|t}^m \, e_2)' \tag{3.3}$$

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

Optimal Memory Loading Matrices: Deriving $\tilde{\Lambda}_t$ The process of determining $\tilde{\Lambda}_t$ resembles Principal Component Analysis (PCA) in that it identifies and prioritizes key information. However, a key distinction lies in the function of the memory system: it *complements* the external data that will be available to the DM in future periods. Unlike PCA, the derivation of $\tilde{\Lambda}_t$ accounts for the interaction with the attention system, as both jointly determine forecast accuracy. This interaction guides which information should be prioritized when the memory system compresses the high-dimensional input space.

Deriving the optimal memory system in the expanded state space presents two main challenges. First, the information to be stored is multi-dimensional, requiring an optimization approach that anticipates future data availability. Second, the loss function is non-convex with respect to the choice variable—not a single matrix, but a sequence of matrices defining the memory system over time. This non-convexity complicates the search for an optimal solution, as it entails global optimization over an infinite sequence of matrices. To make the problem tractable, I simplify the second challenge by considering a myopic case ($\beta \to 0$), which yields an analytical solution for $\tilde{\Lambda}_t$, presented in the following proposition.

Proposition 5. *The matrix* Γ_t *defined as*

$$\Gamma_t \equiv \left(I - \tilde{K}_t\right)' \tilde{Q} \left(I - \tilde{K}_t\right) \tag{3.4}$$

can be eigen-decomposed as follows:

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

where U_t *is an orthonormal matrix containing the eigenvectors, and* G_t *is a diagonal matrix with the eigenvalues arranged in descending order (i.e.,* $g_{1t} > g_{2t}$*). The optimal memory loading matrix* $\tilde{\Lambda}_t$ *is then derived as:*

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

where D_t *is a diagonal matrix of the following form:*

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

Proof. See Appendix A.1 for a full derivation. □

The matrix Γ_t plays a central role in determining which components the memory system should prioritize.[14] This prioritization reflects two main forces: (i) larger values in $I - \tilde{K}_t$ indicate that external data provides little clarity about certain components, increasing reliance on memory; and (ii) the loss function matrix $\tilde{Q}$ weights forecast errors across different dimensions, directing memory towards more consequential sources of error.

Eigen-decomposition guides the memory system to allocate resources efficiently, focusing on directions that minimize forecast errors. The optimization process identifies principal components of Γ_t that are most useful given the DM's prior uncertainty, captured by $Var\left[x_{i,t|t-1}\right]$. These directions guide how memory resources are allocated. The matrix D_t assigns weights to these components in $\tilde{\Lambda}_t$, prioritizing those that explain more variation while respecting the memory constraint ϕ_m.

Two cases illustrate how this plays out. In the **low-rank case**, only the leading principal component is retained: Dt assigns full weight to the first component and zero to the second. This occurs when the memory constraint is tight relative to the ratio, g_{2t}/g_{1t}, where

[14] Γ_t captures the interaction of attention and memory systems. Because the Kalman gain $\tilde{K}_t$ depends on $\tilde{\Lambda}_t$ via its influence on $\Sigma^m_{t|t}$, the memory system indirectly affects itself.

g_{1t} and g_{2t} are the leading eigenvalues of the matrix $Var\left[x_{i,t|t-1}\right]^{1/2}\Gamma_t Var\left[x_{i,t|t-1}\right]^{1/2}$.[15] In the **full-rank case**, both components are retained. D_t assigns more weight to the first component, but the second receives a non-zero allocation. The resulting memory representation remains two-dimensional, reflecting a broader need to track multiple sources of variation.

3.2 Forecast Sensitivity and Stickiness: Short-Term vs. Long-Term

In Section 2.2, I showed that memory frictions increase forecast sensitivity and reduce stickiness when the DM learns about z_t. We now revisit this prediction in the context of long-run learning, examining when and how memory frictions have a stronger influence on belief formation.

The key distinction between short- and long-term learning in the model lies in the availability of external data. I assumed that external data provide only limited information about the persistent latent component, μ_t, making long-term trends harder to infer than near-term fluctuations in z_t. As a result, long-term forecasts rely more heavily on prior knowledge, and memory frictions play a larger role in shaping them.

To investigate this, I revisit the model's predictions for forecast sensitivity and stickiness, focusing on whether memory frictions have a stronger effect on long-term forecasts. I compute forecast outcomes across combinations of attention and memory frictions, and compare them to a rational benchmark with no cognitive frictions. The parameters for this exercise are: $\rho = 0.8$, $\sigma_\epsilon^2 = 1.0$, $\rho_\mu = 0.99$, and $\sigma_\mu^2 = 0.02$ for the data-generating process, with the noise in external data set as $\sigma_\nu^2 = 0.08$. Under this parameterization, the long-run trend accounts for 30 percent of the total variance in z_t, while external noise accounts for 2 percent—typical for the variables analyzed in Section 4.

Figure 2 shows forecast sensitivity for short- and long-term forecasts. The left panel demonstrates that short-term forecasts are generally less responsive to new information than the no-friction benchmark across most of the friction space. Sensitivity exceeds the benchmark only when attention frictions are minimal. In contrast, the right panel shows that long-term forecasts are more responsive to news across a wider range of friction levels. Even with substantial attention frictions, a moderate level of memory friction is sufficient to raise sensitivity above the no-friction benchmark. This heightened responsiveness reflects the limited availability of external data on the long-run trends, which amplifies the influence of memory frictions in shaping forecasts for μ_t.

[15]The eigenvector associated with g_{1t} represents the most informative direction, conditional on prior uncertainty.

Figure 3 presents forecast stickiness for both short- and long-term forecasts. The left panel shows that short-term forecasts are generally stickier than the no-friction benchmark across much of the friction space. Stickiness falls below the benchmark only when attention frictions are very low or memory frictions are high. In contrast, the right panel shows that long-term forecasts exhibit reduced stickiness across a broader range of friction levels. Even with substantial attention frictions, moderate memory frictions are enough to lower stickiness below the benchmark. This effect occurs because memory frictions weaken reliance on prior beliefs, causing long-term forecasts to become less anchored to the past.

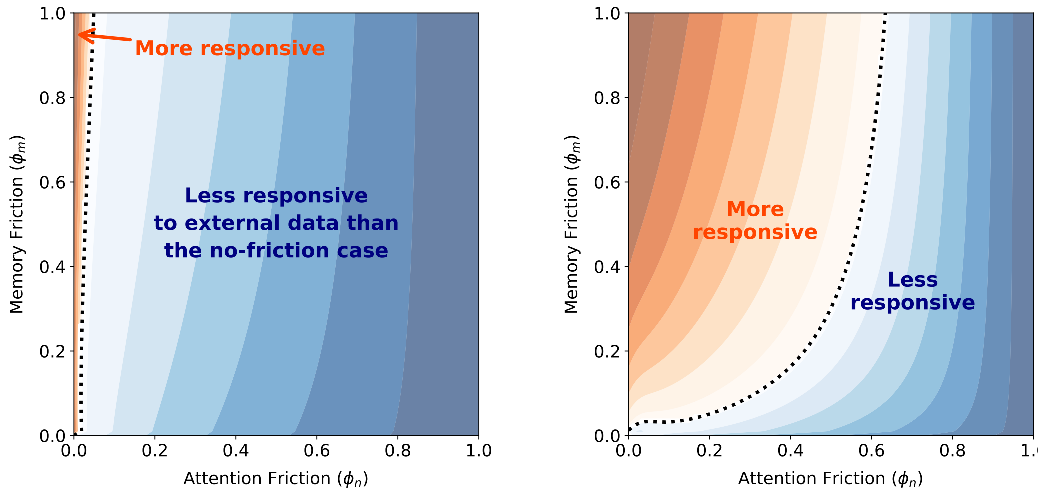
Forecast Bias Across Horizons Figures 2-3 suggest that spike-and-decay dynamics become increasingly prominent at longer forecast horizons. In Appendix Section E, I explore how similar patterns emerge when long-run uncertainty arises from parameter uncertainty rather than stochastic fluctuations. Revisiting the forecast error-revision regressions (2.11) and (2.10), I also show that the associated regression coefficients become increasingly negative as the forecast horizon lengthens. Additionally, I discuss how the heightened sensitivity of long-term forecasts offers an explanation for empirical patterns documented in Angeletos et al. (2021), who document that professional forecasters' year-ahead predictions for unemployment and inflation initially undershoot but eventually overshoot realized outcomes.[16]

This horizon-dependent bias pattern has been documented in various contexts. Halperin and Mazlish (2025) find positive regression coefficients for near-term forecasts and increasingly negative coefficients for longer-term forecasts across GDP, inflation, investment and consumption for multiple countries. Afrouzi et al. (2023) find similar patterns in laboratory experiments, while Bordalo et al. (2019) and Bordalo et al. (2023) report analogous results for stock analysts' forecasts of long-term corporate earnings. Similarly, both d'Arienzo (2020) and Wang (2021) find that forecasts for longer-maturity interest rates exhibit increasingly negative biases when estimating (2.11) and (2.10).

[16] Specifically, Angeletos et al. (2021) analyze impulse responses of professional forecasts to shocks constructed by Angeletos et al. (2020).

Figure 2: Forecast Sensitivity: Short-term vs. Long-term

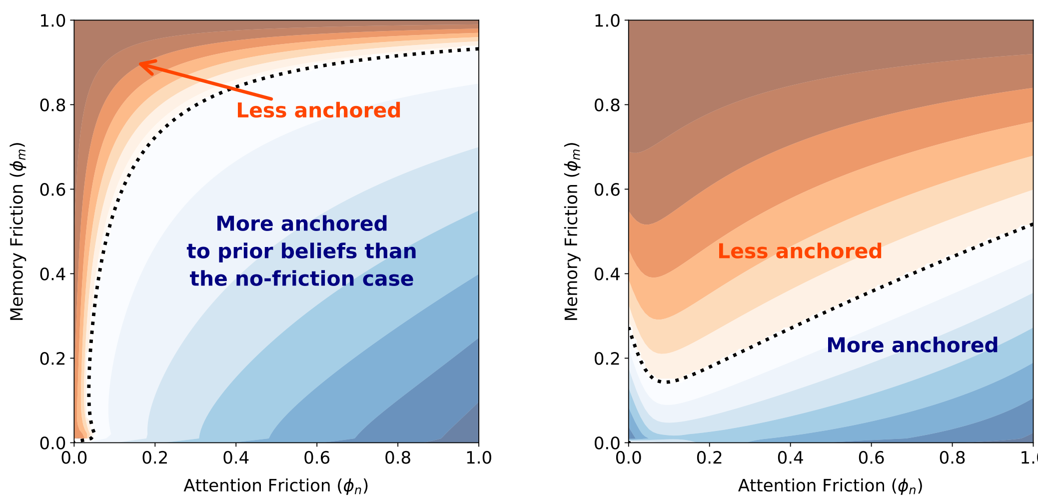
(a) Forecasts of z_t (b) Forecasts of μ_t



The figure shows forecast sensitivity to new information for short-term (left panel) and long-term (right panel) forecasts, relative to a rational benchmark with no cognitive friction, across varying levels of cognitive frictions. The data-generating process underlying these results is described in detail in the main text.

Figure 3: Forecast Stickiness: Short-term vs. Long-term

(a) Forecasts of z_t (b) Forecasts of μ_t



The figure shows forecast stickiness for short-term (left panel) and long-term (right panel) forecasts, relative to a rational benchmark of no cognitive friction, across varying levels of cognitive frictions. The data-generating process underlying these results is described in detail in the main text.

4 Estimating the Magnitude of Cognitive Frictions

This section brings the model to the data by estimating the magnitude of attention and memory frictions using survey forecasts. Quantifying these frictions allows the model to match key features of forecast behavior and sets up the policy analysis in the next section.

4.1 Data

The survey forecast data come from the Survey of Professional Forecasters (SPF), administered by the Federal Reserve Bank of Philadelphia. Conducted quarterly since 1968, the survey typically includes responses from approximately forty forecasters. I use forecasts made through the fourth quarter of 2024.

The estimation focuses on inflation, output, and interest rate forecasts. Specifically, I examine three measures of inflation (the year-over-year percentage change in CPI, PCE, and the chain-weighted GDP price index), two measures of output (the year-over-year percentage change in nominal GDP and real GDP), and four measures of interest rates (the quarterly average of the Treasury bill rate, the 10-year Treasury bond rate, and Moody's Aaa and Baa corporate bond yields).

To construct the time series of realized macroeconomic outcomes, I use the Real-Time Data Set from the Federal Reserve Bank of Philadelphia, which records the full revision history of each variable.[17] I use the vintage available three quarters after the reference quarter, substituting earlier releases when necessary.[18]

4.2 Estimation Strategy

The estimation proceeds in two parts: I first estimate empirical moments that serve as targets that the model's friction parameters are chosen to match, and then calibrate parameters that shape the environment in which those frictions operate.

Estimation targets. My primary goal is to quantify attention and memory frictions that influence forecasters' expectations. I estimate these frictions by targeting empirical regres-

[17]Vintage data are particularly important for variables from the National Income and Product Accounts (NIPA), which are often redefined or reclassified. As a result, the most recent data may not reflect the definitions used by forecasters. The earliest release ("Advance" estimate) is typically published about one month after the quarter ends, with subsequent releases incorporating additional information.

[18]This choice balances two goals: using data that are sufficiently revised to reduce measurement error, while avoiding later revisions that may reflect definitional changes. In practice, using this vintage versus the first release makes little difference for the analyses that follow.

sion coefficients that capture systematic patterns of forecast errors. Specifically, I focus on two regression-based moments: β_C of regression (2.11) and β_I of regression (2.10).

For the estimation of β_C, I adopt a modified version of regression (2.11) to address several theoretical and statistical issues with the original formulation proposed by Coibion and Gorodnichenko (2015).[19] Following Goldstein (2023) and Gemmi and Valchev (2023), I estimate the following demeaned version:

$$F_{i,t}\, y_{t+h} - F_t\, y_{t+h} = \alpha_C + \beta_C \left(F_{i,t-1}\, y_{t+h} - F_{t-1}\, y_{t+h}\right) + error_{i,t+h|t-1} \tag{4.1}$$

This specification captures the persistence of individual deviations from the consensus forecast. Intuitively, when relative forecast positions are highly persistent, updates are limited—implying low Kalman gains. Compared to the original formulation, this version removes the common noise component ($\tilde{\nu}_t$ in the model), allowing for cleaner identification of individual stickiness.[20]

To improve statistical power, I estimate each regression on the full panel of forecasters, including forecaster-specific intercepts to account for differences in average levels. Thus, this approach imposes a common slope across individuals, while allowing for heterogeneity in intercepts. Standard errors are clustered two ways, by forecaster and by time period, to account for both serial and cross-sectional dependence. I also use the Huber robust estimator to mitigate the influence of outliers.

To focus on expectation formation under typical business cycle conditions, I exclude periods of unusually large shocks or structural changes, during which the estimated statistical relationships in (4.1) and (2.10) appear less stable. Specifically, I drop the 10% of survey periods with the highest mean squared forecast errors, measured as the average squared gap between consensus forecasts and realizations in each period.

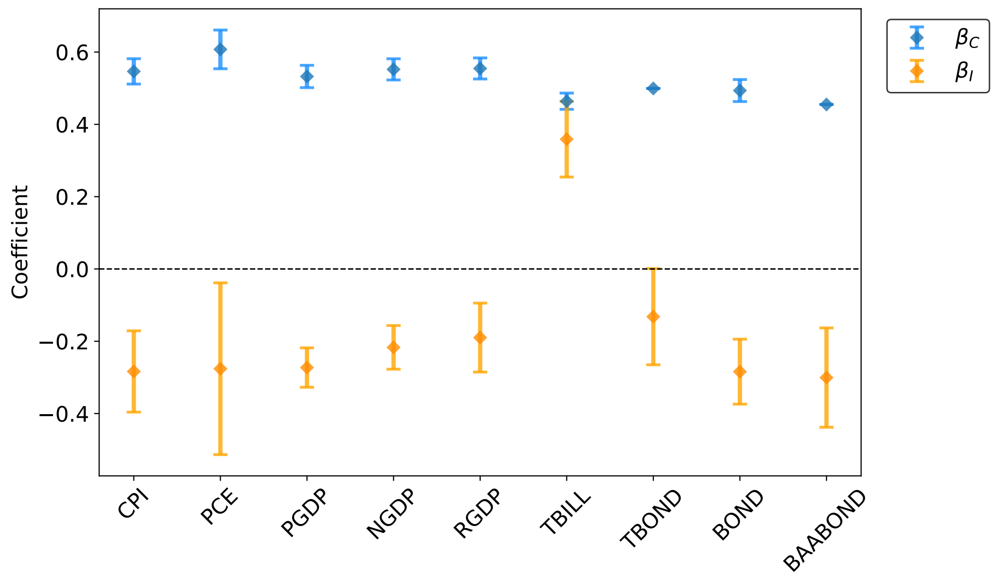
For all macroeconomic variables, I use three-quarter-ahead forecasts. Figure 4 presents the resulting estimates. As discussed in Section 2.3, β_C is generally positive, while β_I tends to be negative. The Treasury bill rate is an exception, with a positive β_I likely driven by factors unrelated to memory frictions. Therefore, I estimate the cognitive frictions using the remaining eight variables.

Specification of remaining parameters. To complete the estimation, several model parameters must be calibrated.

[19] As noted by the authors, the regression coefficient is attenuated when forecasters' signal noise is correlated. Other issues include the unreliability of estimates under structural breaks (Hajdini & Kurmann, 2023) and limited out-of-sample predictability (Bianchi et al., 2022). These concerns suggest that the original specification is subject to small-sample problems.

[20] I retain the notation β_C to reduce notational burden.

Figure 4: Regression Coefficients for Estimation Targets



The figure reports the regression coefficients used as estimation targets. The top panel displays the coefficient from regression (4.1), and the bottom panel shows those from (2.10). For both, I use Huber-robust estimation with individual fixed effects and two-way clustered standard errors (by forecaster and date). Table 1 lists the regression coefficients shown in the figure.

First, I estimate the data-generating process of macroeconomic variables (equation 3.1) by decomposing each series into three frequency bands using the approximate bandpass filter of Baxter and King (1999).[21] The low-frequency trend represents the persistent latent state (μ_t), and the medium-frequency cycle captures transitory fluctuations ($z_t - \mu_t$). The high-frequency component, reflecting short-lived movements irrelevant at a three-quarter horizon, is treated as external signal noise ($\tilde{\nu}_t$ in equation 2.4). I calibrate the variance of this noise to match the estimated variance of the irregular component. Appendix Table 2 reports the full estimates.

Second, I assume that forecasters consider horizons up to eight quarters ahead in their loss function (1.1), consistent with the SPF's two-year forecast horizon. Results remain stable when the horizon is extended.

4.3 Estimation Results

Figure 5 displays the estimated attention frictions (ϕ_n) and memory frictions (ϕ_m) parameters, accompanied by their two-standard-deviation confidence intervals. The top panel illustrates the estimates for ϕ_n, while the bottom panel depicts the estimates for ϕ_m. Im-

[21] The trend cutoff is at 32 quarters; the irregular cutoff at 3 quarters, with 12 quarters of leads and lags.

portantly, the estimates shown here exactly reproduce the empirical targets presented in Figure 4, although such a precise match is not mechanically ensured.

My estimates of attention frictions (ϕ_n) are approximately twice as large as those implied by conventional approaches. In the top panel, blue squares represent estimates from my method, while gray triangles are estimates from standard noisy information models that assume perfect memory. As explained in Section 2.3.4, this discrepancy arises because conventional methods—such as Coibion and Gorodnichenko (2015)—misattribute the heightened sensitivity caused by memory frictions to lower attention frictions.

Figure 6 assesses the model's ability to match non-targeted empirical moments. Specifically, I examine whether the estimated model reproduces cross-sectional dispersion observed in survey forecasts. For each variable, I report the median, 2.5th, and the 97.5th percentiles of forecast dispersion, controlling for individual fixed effects. These empirical moments are then compared to the model's predictions for the same measures. While the survey data exhibit substantial fluctuations in dispersion over time, the model captures the typical magnitude of this variation through cognitive noise, except for Moody's Aaa and Baa corporate bond yields, where the model underpredicts the dispersion.

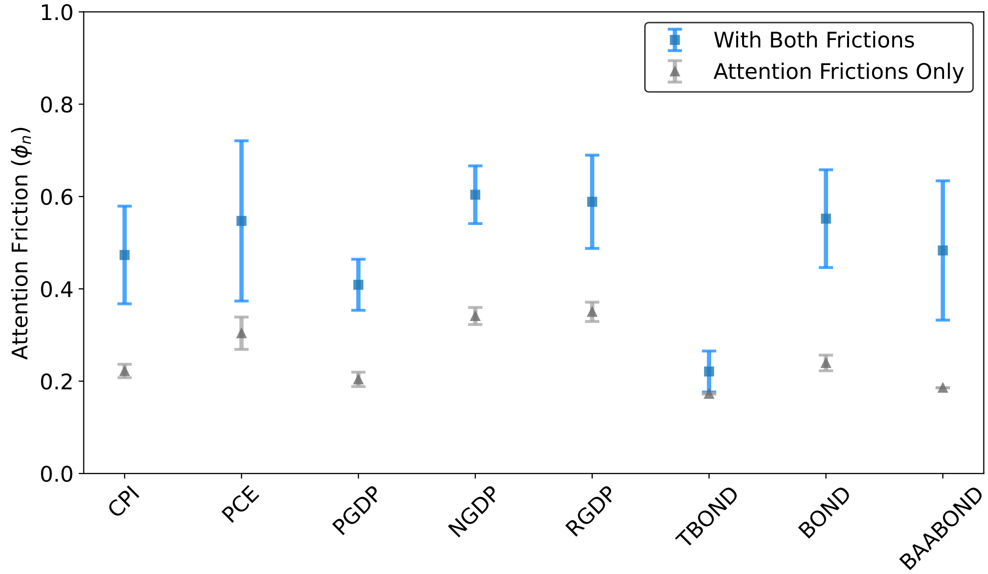
Beyond capturing cross-sectional variations, the estimated cognitive frictions shed light on how expectations evolve over time. Specifically, they help us understand how forecasts for low-frequency latent variables respond dynamically to economic shocks—an important insight given the limited availability of long-term survey data. Figure 7 shows the model-implied impulse response functions for PCE inflation expectations. The solid gray line is the response of the transitory inflation component (z_t) to a shock. The dashed orange lines trace the evolution of expectations under the estimated attention and memory frictions, while the dotted black lines represent the frictionless benchmark. Analogous results for other macroeconomic variables are presented in Appendix G.

Without cognitive frictions, short-term forecasts of the transitory component (z_t) closely track actual realizations. However, long-term forecasts of the persistent component (μ_t) remain elevated even 20 quarters after the initial shock—beyond the point at which its effects on fundamentals have largely dissipated.

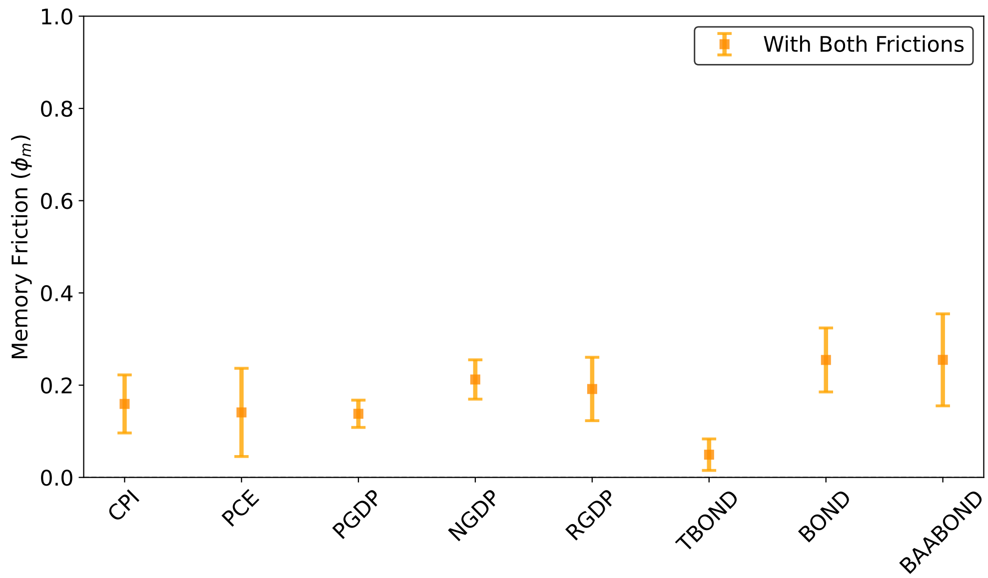
The response under estimated attention and memory frictions aligns with the model's predictions in Section 3.2. Short-term forecasts initially underreact, with gradual updating reducing the gap over time. Long-term forecasts also start with underreaction, but briefly overshoot before converging toward steady-state levels as memory decays. This pattern reflects the spike-and-fade dynamics of memory frictions: they first amplify forecast sensitivity to new information, then reduce stickiness by accelerating reversion to steady-state.

Figure 5: Estimation Results on Cognitive Frictions

(a) Attention Frictions

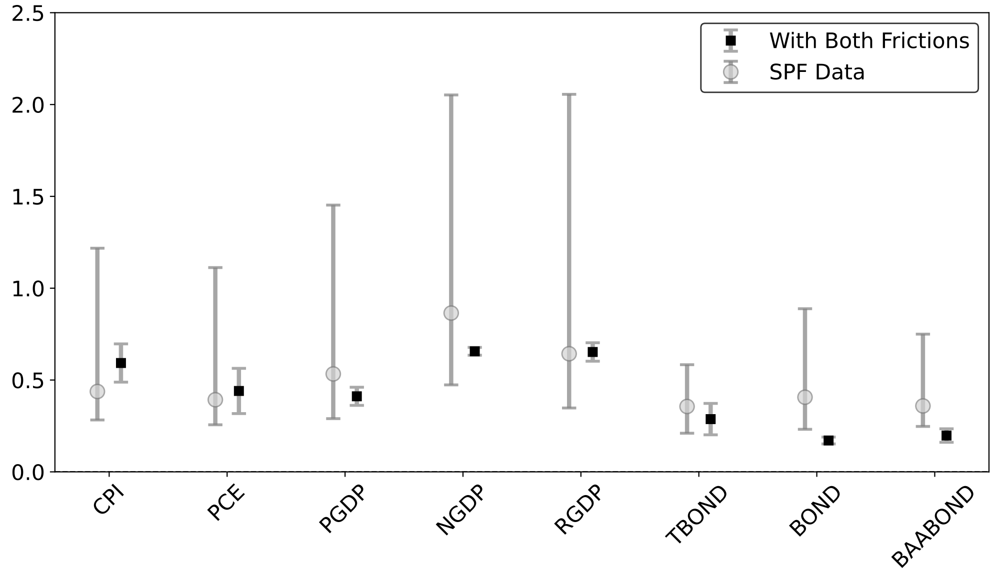


(b) Memory Frictions



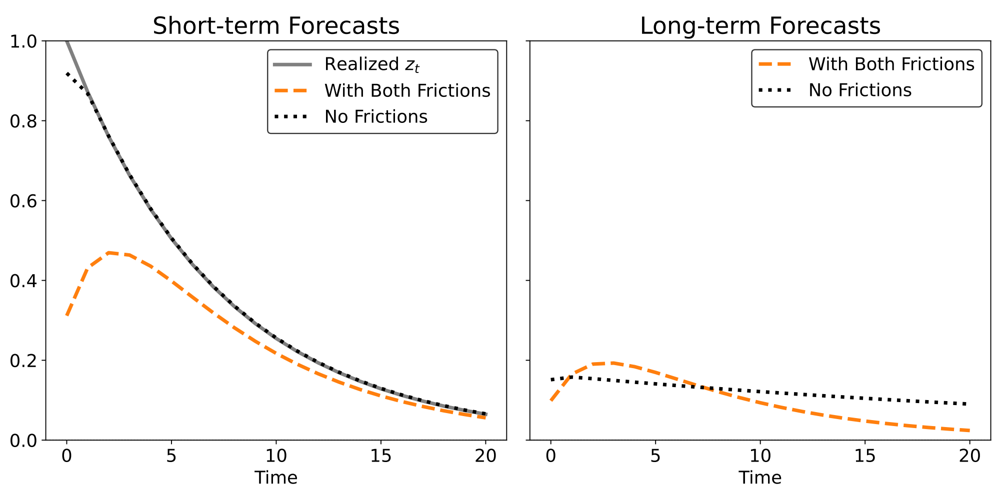
The figure reports the estimated cognitive frictions. The top panel shows the estimated attention frictions (ϕ_n) and the two-standard-deviation confidence intervals. The blue lines correspond to the proposed estimation, while the gray lines are estimates from traditional models of information frictions that assume perfect memory. The bottom panel displays the estimated memory frictions (ϕ_m). Table 3 reports the estimated magnitude of frictions in the figure.

Figure 6: Estimated Model's Prediction of the Cross-sectional Variations



The figure tests the model fit of non-targeted moments. The circle markers show the cross-sectional variations in the survey data and the error bars are the 2.5 and 97.5 percentiles, while controlling for forecaster fixed effects. The square markers are the model predictions of the same object with the error bars depicting the two standard deviations. Table 4 provides the numerical values underlying the figure.

Figure 7: Estimated Impulse Response of Inflation Expectations (PCE)



The figure displays the impulse responses of forecasts about the latent drivers of PCE inflation. Solid gray line shows the response of the latent variable. Dashed orange lines represent forecasts under estimated cognitive frictions, while dotted black lines represent the frictionless benchmark. The left panel shows forecasts of the short-term component (z_t), and the right panel shows forecasts of the long-term component (μ_t).

5 Implications for Monetary Policy

The estimated expectation dynamics for PCE inflation from the previous section serves as a key input for analyzing optimal monetary policy responses to cost-push shocks. By explicitly integrating these empirically grounded expectations into the model, the analysis highlights how cognitive frictions alter the policy trade-offs faced by monetary authorities.

5.1 Firms' Optimal Price Setting

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

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

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 demand that firm i faces at $t+h$ if its price remains at the level set in t, and Ψ_{t+h} denotes nominal costs at time $t+h$. Demand is given by:

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

where η is the elasticity of substitution across goods, P_{t+k} is the aggregate price, and C_{t+h} is the aggregate consumption. I use the notation $E_{i,t}$ to denote firm i's *subjective* expectation, departing from the conventional full-information rational expectations assumption.

The first-order condition characterizing the firm's optimal price is log-linearized around the zero-inflation steady state as:

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

where lowercase variables denote logs, mc_{t+h} is the log of real marginal cost, mc its steady-state value, and π_{t+h} is inflation at $t+h$ defined as $\log P_{t+h} - \log P_{t+h-1}$. β is the time discount factor.

Let z_t denote the nominal marginal cost component relevant for the firm's pricing decision:

$$z_t \equiv (1-\alpha\beta)\left(mc_t - mc\right) + \pi_t. \tag{5.1}$$

The firm's optimal pricing decision is then governed by:

$$p_{i,t}^{*} - p_{i,t-1} = E_{i,t}\left[\sum_{h=0}^{\infty} (\alpha\beta)^{h} z_{t+h}\right] \tag{5.2}$$

I assume that firms treat z_t as independent of their own price choices, viewing it as a function of aggregate variables alone.[22]

5.2 Aggregate Economy

Price changes at the firm level aggregate to the overall inflation rate. The aggregate price index evolves according to:

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

where $P_t^{*} \equiv \int P_{i,t}^{*}\, di$ is the average reset price. A first-order approximation yields

$$\pi_t = (1-\alpha)\left(p_t^{*} - p_{t-1}\right)$$

and substituting the firm's pricing rule gives:

$$\pi_t = (1-\alpha)\bar{E}_t\left[\sum_{h=0}^{\infty} (\alpha\beta)^{h} z_{t+h}\right] \tag{5.3}$$

where $\bar{E}_t$ is the average of all firms' expectations.

Real marginal costs are pinned down by household optimization and market clearing. As shown in Appendix H,

$$mc_t - mc = (\sigma + \varphi)\,\tilde{y}_t + e_t. \tag{5.4}$$

where σ is the intertemporal elasticity of substitution and φ is the inverse Frisch elasticity of labor supply, and $\tilde{y}_t$ is the output gap, defined as $y_t - y_t^{e}$, where y_t^{e} is the efficient level of output. Finally, e_t is the cost-push shock.

Substituting the expression for $mc_t - mc$ into the definition of z_t in equation (5.1) yields:

$$z_t = \pi_t + (1-\alpha\beta)(\sigma+\varphi)\,\tilde{y}_t + (1-\alpha\beta)\,e_t. \tag{5.5}$$

Thus, the joint evolution of inflation, output, and marginal costs is governed by how firms forecast the path of z_t.

[22]As detailed in Appendix H, marginal costs do not depend on a firm's individual output due to properties of the assumed production function. In this case, subjectively optimal pricing still depends only on beliefs about aggregate conditions. See Gali (2008, Ch. 3).

5.3 Firms' Inflation Expectations

I assume firms form expectations of current and future values of z_t according to the expectation models for PCE inflation estimated in Section 4. In doing so, I treat firms' beliefs about the underlying inflation process—such as the law of motion governing the latent inflation factors—as fixed. These beliefs are parameterized by coefficients derived from observed inflation dynamics and held constant throughout the analysis.

Although firms apply a fixed forecasting model, their expectations do respond to monetary policy, since the latent factors used in the model are influenced by policy and aggregate disturbances. However, firms fail to internalize how the structure of inflation dynamics might shift following changes in policy regimes.[23] This modeling choice thus represents a partial equilibrium approach, best interpreted as capturing short-term policy responses in an economy where firms' perceived inflation dynamics adjust slowly.

Despite this simplification, the expectations formation process retains several realistic features. Firms distinguish between persistent and transitory cost pressures and form expectations subject to attention and memory frictions, with the level of each estimated from survey data. These frictions influence the magnitude and persistence of inflation responses, offering insights into how monetary policy should be designed under cognitive constraints.

5.4 Parameterization

I parameterize the model using conventional values: $\sigma = \varphi = 1$ (log utility and log disutility of labor), $\beta = 0.99$ (quarterly discounting), and $\alpha = 2/3$ (average price duration of three quarters). The cost-push shock follows an AR(1) process with a persistence of 0.8.

[23] This internal inconsistency reflects the need to balance realism and tractability of the model. The structural model employed here is deliberately simplified to transparently illustrate the key monetary policy trade-offs associated with cost-push shocks. Although the model captures crucial aspects of how firms' expectations and marginal costs respond to output gaps and shocks, it necessarily omits many realistic complexities of actual inflation dynamics. A fully realistic, equilibrium-consistent expectations formation process would significantly complicate the analysis.

5.5 Optimal Monetary Policy

The central bank minimizes a quadratic loss function representing a dual mandate with equal weights on inflation and output-gap stabilization:[24]

$$\min_{\{\pi_t, \tilde{y}_t\}_t} E \sum_{t=0}^{\infty} \beta^t \left(\pi_t^2 + \tilde{y}_t^2\right) \tag{5.6}$$

subject to the inflation dynamics and pricing behavior described above, i.e. equations (5.5) and (5.3). I assume the central bank correctly understands the structure of the economy, including how firms form expectations.

Figure 8 shows how the optimal policy response to cost-push shocks varies with the severity of cognitive frictions. In both panels, inflation rises and output contracts as the central bank dampens aggregate demand to contain price pressures.

Panel (a) shows the effects of varying attention frictions. When attention frictions are more severe, firms process information about cost conditions less accurately. As a result, their prices adjust more sluggishly in response to shocks. This muted price response dampens the inflationary impact of the shock, allowing the central bank to stabilize inflation with a milder contraction in output.

Panel (b) highlights the more nuanced effects of memory frictions. When memory frictions are stronger, firms respond more to shocks in the short run, but their expectations fade quickly over time. This creates an ambiguous policy trade-off. On the one hand, the central bank might consider tightening early to counteract the sharper initial response of firms. On the other hand, the natural decay of expectations reduces the need for aggressive intervention, since inflationary pressures will dissipate on their own. The model shows that the latter effect dominates: the central bank leans toward a "look-through" stance, tolerating a temporarily higher path of inflation in order to avoid an unnecessarily large contraction in output.

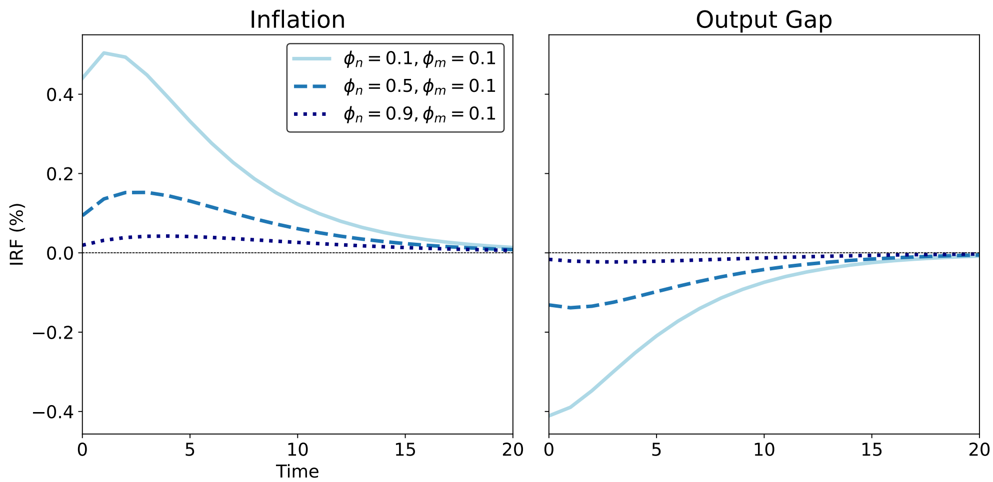
Although both weaker attention frictions and stronger memory frictions increase the short-run responsiveness of inflation, they lead to opposite policy prescriptions. When firms process shocks more accurately, the central bank responds more forcefully to prevent inflation from drifting upward, resulting in a deeper output contraction. In contrast, when expectations are short-lived due to memory frictions, inflation pressures fade on their own. The central bank can therefore tolerate temporarily higher inflation and tighten less, lead-

[24]This objective is not explicitly derived from the welfare of the representative consumer. Deriving a welfare-based loss function under the expectation formation subject to cognitive constraints proposed in this paper would provide new insights into how deviations from rational expectations alter the standard loss function. However, my primary focus is to isolate how optimal policy differs across various expectation assumptions under a consistent policy objective; thus, I leave this extension for future research.

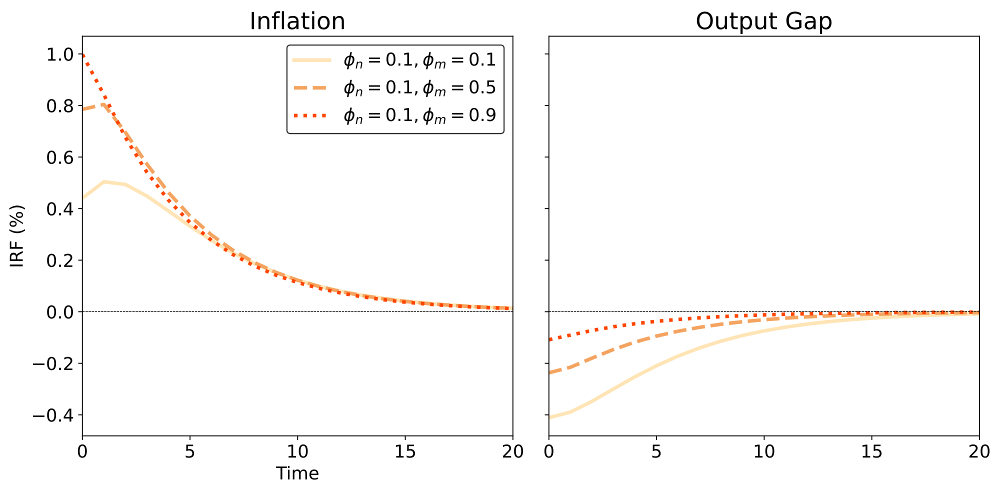
ing to a milder contraction in output.

Figure 8: Optimal Monetary Policy under Varying Cognitive Frictions

(a) Variation in Attention Frictions



(b) Variation in Memory Frictions

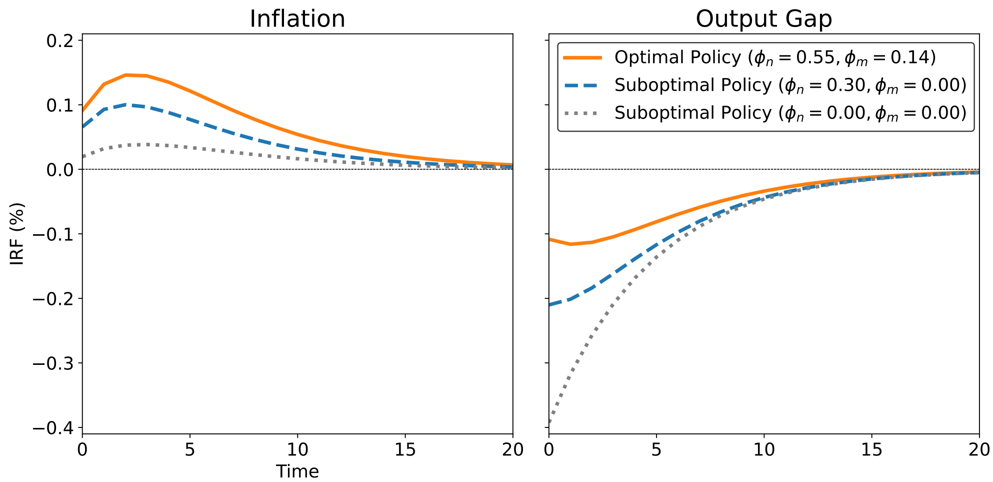


The figure plots the optimal path of inflation and the output gap in response to a cost-push shock. Panel (a) shows how optimal policy responds to varying degrees of attention frictions, while panel (b) illustrates the effects of memory frictions.

Finally, Figure 9 compares inflation and output dynamics under different monetary policy assumptions. The orange line shows the optimal policy that accounts for the estimated levels of both attention and memory frictions. The dashed blue line applies a policy optimized under attention frictions alone, using the estimates from Section 4 that assume perfect memory. This policy overestimates the responsiveness of inflation and tightens too aggressively, resulting in larger output losses. The dotted gray line represents policy under full-information rational expectations. Here, the central bank responds even more forcefully to the shock, further amplifying the output loss.

These comparisons make clear that policy prescriptions are sensitive to how expectations are modeled. In this setting, omitting memory frictions leads to overly aggressive stabilization and unnecessary real costs. More broadly, basing policy decisions on empirically realistic models of expectation formation leads to more appropriate policy design and improves our ability to predict the dynamics of macroeconomic variables.

Figure 9: Counterfactual Monetary Policy



The figure compares inflation and output dynamics in response to a cost-push shock under three different policy assumptions. Each line corresponds to the optimal policy under a different expectation model: (1) the baseline case with both attention and memory frictions, (2) a model with attention frictions only (assuming perfect memory), and (3) full-information rational expectations with no frictions. The values of the frictions are estimated for the PCE inflation, as described in Section 4.

6 Conclusion

This paper extends the standard models of information frictions by introducing memory frictions alongside attention frictions. Memory frictions produce distinctive "spike-and-fade" dynamics, where expectations respond sharply to news but fade quickly over time. Incorporating both frictions allows the model to reconcile conflicting empirical findings, such as sluggish adjustments in short-term forecasts and the excess sensitivity of long-term forecasts. Structural estimation using survey forecast data shows that attention frictions are substantially underestimated when memory frictions are ignored. The estimated friction parameters suggest that monetary policy should tolerate cost-push shocks more, since memory frictions cause elevated long-term inflation expectations to fade more quickly.

More broadly, these results underscore the importance of cognitive constraints in shaping expectations and guiding optimal policy.

7 Additional Tables

Table 1: Regression Coefficients for Estimation Targets

	β_C	SD	β_I	SD
CPI Inflation	0.55	0.02	-0.28	0.06
PCE Inflation	0.61	0.03	-0.28	0.12
PGDP Inflation	0.53	0.02	-0.27	0.03
Nominal GDP Growth	0.55	0.01	-0.22	0.03
Real GDP Growth	0.55	0.01	-0.19	0.05
3-Month T-Bill Rate	0.46	0.01	0.36	0.05
10-Year T-Bond Rate	0.50	0.00	-0.13	0.07
Aaa Corporate Bond Yields	0.49	0.02	-0.28	0.05
Baa Corporate Bond Yields	0.45	0.00	-0.30	0.07

This table reports the regression coefficients used as estimation targets in Figure 5. The first two columns show estimates from regression (4.1); the last two columns report estimates from regression (2.10). All regressions include individual fixed effects and are estimated using the Huber robust estimator. Standard errors are two-way clustered by forecaster and date.

Table 2: Data-generating Process

	ρ	σ_ϵ	μ	ρ_μ	σ_μ	σ_ν
CPI Inflation	0.84	0.40	6.07	0.99	0.11	0.10
PCE Inflation	0.85	0.58	4.34	0.99	0.14	0.15
PGDP Inflation	0.82	1.34	2.80	0.98	0.11	0.50
Nominal GDP Growth	0.88	0.41	3.19	0.99	0.10	0.14
Real GDP Growth	0.87	0.50	3.08	0.99	0.11	0.16
3-Month T-Bill Rate	0.80	1.38	6.06	0.99	0.12	0.54
10-Year T-Bond Rate	0.89	0.58	3.50	0.99	0.14	0.18
Aaa Corporate Bond Yields	0.85	0.25	6.18	0.99	0.03	0.07
Baa Corporate Bond Yields	0.85	0.27	6.80	0.99	0.05	0.09

This table reports estimated parameters of the data-generating processes for the macroeconomic variables used in the model. For each series, ρ and σ_ϵ describe the AR(1) process for the cycle component, while μ, ρ_μ, and σ_μ characterize the trend component. The final column, σ_ν, reports the standard deviation of external signal noise, calibrated based on standard deviation of the irregular component.

Table 3: Estimation Results on Cognitive Frictions

	Baseline				Attention Frictions Only			
	ϕ_n	SD	ϕ_m	SD	ϕ_n	SD	ϕ_m	SD
CPI Inflation	0.47	0.11	0.16	0.06	0.22	0.01	0.00	-
PCE Inflation	0.55	0.17	0.14	0.10	0.30	0.04	0.00	-
PGDP Inflation	0.41	0.06	0.14	0.03	0.20	0.02	0.00	-
Nominal GDP Growth	0.60	0.06	0.21	0.04	0.34	0.02	0.00	-
Real GDP Growth	0.59	0.10	0.19	0.07	0.35	0.02	0.00	-
10-Year T-Bond Rate	0.22	0.04	0.05	0.03	0.17	0.00	0.00	-
Aaa Corporate Bond Yields	0.55	0.11	0.25	0.07	0.24	0.02	0.00	-
Baa Corporate Bond Yields	0.48	0.15	0.25	0.10	0.19	0.00	0.00	-

This table reports the estimated magnitude of cognitive frictions shown in Figure 5. The columns display attention frictions (ϕ_n) and memory frictions (ϕ_m), along with two-standard deviation standard errors, under two estimation approaches. The "Baseline" columns correspond to the proposed model, which incorporates both attention and memory frictions. The "Attention Frictions Only" columns correspond to a restricted model that assumes perfect memory.

Table 4: Estimated Model's Prediction of the Cross-sectional Variations

	Data (SPF)			Model	
	2.5%	50%	97.5%	Mean	SD
CPI Inflation	0.28	0.44	1.22	0.59	0.10
PCE Inflation	0.26	0.39	1.11	0.44	0.12
PGDP Inflation	0.29	0.53	1.45	0.41	0.05
Nominal GDP Growth	0.47	0.86	2.05	0.66	0.02
Real GDP Growth	0.35	0.64	2.06	0.65	0.05
10-Year T-Bond Rate	0.21	0.36	0.58	0.29	0.09
Aaa Corporate Bond Yields	0.23	0.41	0.89	0.17	0.02
Baa Corporate Bond Yields	0.25	0.36	0.75	0.20	0.04

This table reports the numerical values underlying Figure 6, which evaluates model fit for non-targeted moments. The "Data (SPF)" columns summarize the cross-sectional distribution of individual forecasts from the Survey of Professional Forecasters, controlling for forecaster fixed effects. The 2.5%, 50%, and 97.5% columns correspond to empirical percentiles of forecast dispersion. The "Model" columns report the model-implied mean and standard deviation of the dispersion across agents.

References

Afrouzi, H., & Yang, C. (2021). *Dynamic Rational Inattention and the Phillips Curve*, Working Paper. CESifo.

Afrouzi, H., Kwon, S. Y., Landier, A., Ma, Y., & Thesmar, D. (2023). Overreaction in Expectations: Evidence and Theory. *Quarterly Journal of Economics*, *138*(3).

Andrade, P., & Le Bihan, H. (2013). Inattentive professional forecasters. *Journal of Monetary Economics*, *60*(8), 967–982.

Angeletos, G.-M., & Huo, Z. (2021). Myopia and Anchoring. *American Economic Review*, *111*(4), 1166–1200.

Angeletos, G.-M., Collard, F., & Dellas, H. (2020). Business-Cycle Anatomy. *American Economic Review*, *110*(10), 3030–3070.

Angeletos, G.-M., Huo, Z., & Sastry, K. A. (2021). Imperfect Macroeconomic Expectations: Evidence and Theory. *NBER Macroeconomics Annual*, *35*(1), 2–86.

Azeredo da Silveira, R., & Woodford, M. (2019). Noisy memory and over-reaction to news. *AEA Papers and Proceedings*, *109*, 557–61.

Azeredo da Silveira, R., Sung, Y., & Woodford, M. (2024). Optimally imprecise memory and biased forecasts. *American Economic Review*, *114*(10), 3075–3118.

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

Bianchi, F., Ludvigson, S. C., & Ma, S. (2022). Belief Distortions and Macroeconomic Fluctuations. *American Economic Review*, *112*(7), 2269–2315.

Bordalo, P., Gennaioli, N., Porta, R. R., & Shleifer, A. (2019). Diagnostic Expectations and Stock Returns. *The Journal of Finance*, *74*(6), 2839–2874.

Bordalo, P., Gennaioli, N., Ma, Y., & Shleifer, A. (2020). Overreaction in Macroeconomic Expectations. *American Economic Review*, *110*(9), 2748–2782.

Bordalo, P., Conlon, J. J., Gennaioli, N., Kwon, S. Y., & Shleifer, A. (2022). Memory and Probability. *The Quarterly Journal of Economics*, *138*(1), 265–311.

Bordalo, P., Gennaioli, N., Porta, R. L., & Shleifer, A. (2023). Belief Overreaction and Stock Market Puzzles. *Journal of Political Economy*.

Bordalo, P., Burro, G., Coffman, K., Gennaioli, N., & Shleifer, A. (2024). Imagining the future: Memory, simulation, and beliefs. *The Review of Economic Studies*, rdae070.

Broer, T., & Kohlhas, A. N. (2022). Forecaster (Mis-)Behavior. *The Review of Economics and Statistics*, 1–45.

Carvalho, C., Eusepi, S., Moench, E., & Preston, B. (2023). Anchored Inflation Expectations. *American Economic Journal: Macroeconomics*, *15*(1), 1–47.

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

Coibion, O., & Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, *120*(1), 116–159.

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

Coibion, O., Gorodnichenko, Y., & Kumar, S. (2018). How Do Firms Form Their Expectations? New Survey Evidence. *American Economic Review*, *108*(9), 2671–2713.

Coibion, O., Gorodnichenko, Y., & Weber, M. (2022). Monetary Policy Communications and Their Effects on Household Inflation Expectations. *Journal of Political Economy*, *130*(6), 1537–1584.

Collin-Dufresne, P., Johannes, M., & Lochstoer, L. A. (2016). Parameter Learning in General Equilibrium: The Asset Pricing Implications. *American Economic Review*, *106*(3), 664–98.

Cover, T. M., & Thomas, J. A. (2005). *Elements of Information Theory*. Wiley-Interscience.

Crump, R. K., Eusepi, S., Moench, E., & Preston, B. (2023). The term structure of expectations. *Handbook of Economic Expectations*, 507–540.

d'Arienzo, D. (2020). Maturity Increasing Overreaction and Bond Market Puzzles. *SSRN Electronic Journal*.

Elliott, G., Komunjer, I., & Timmermann, A. (2008). Biases in Macroeconomic Forecasts: Irrationality or Asymmetric Loss? *Journal of the European Economic Association*, *6*(1), 122–157.

Enke, B., & Graeber, T. (2023). Cognitive Uncertainty. *The Quarterly Journal of Economics*, *138*(4), 2021–2067.

Enke, B., Schwerter, F., & Zimmermann, F. (2024). Associative memory, beliefs and market interactions. *Journal of Financial Economics*, *157*, 103853.

Evans, G. W., & Honkapohja, S. (2012). *Learning and expectations in macroeconomics*.

Farmer, L., Nakamura, E., & Steinsson, J. (2024). Learning About the Long Run. *Journal of Political Economy*.

Fuster, A., Laibson, D., & Mendel, B. (2010). Natural Expectations and Macroeconomic Fluctuations. *Journal of Economic Perspectives*, *24*(4), 67–84.

Gabaix, X. (2014). A sparsity-based model of bounded rationality *. *The Quarterly Journal of Economics*, *129*(4), 1661–1710.

Galí, J. (2015). *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition*. Princeton University Press.

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

Gemmi, L., & Valchev, R. (2023). Biased Surveys. *National Bureau of Economic Research, Working Paper Series*(31607).

Giglio, S., & Kelly, B. (2018). Excess Volatility: Beyond Discount Rates. *The Quarterly Journal of Economics*, *133*(1), 71–127.

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

Hajdini, I., & Kurmann, A. (2023). *Predictable Forecast Errors in Full-Information Rational Expectations Models with Regime Shifts*, Working Paper. Federal Reserve Bank of Cleveland.

Halperin, B., & Mazlish, J. Z. (2025, April). *Overreaction and Forecast Horizon: Longer-term Expectations Overreact More, Shorter-term Expectations Drive Fluctuations* (Economics Series Working Papers No. 1076), University of Oxford, Department of Economics.

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

Hazell, J., Herreño, J., Nakamura, E., & Steinsson, J. (2022). The Slope of the Phillips Curve: Evidence from U.S. States. *The Quarterly Journal of Economics*, *137*(3), 1299–1344.

Juodis, A., & Kučinskas, S. (2023). Quantifying noise in survey expectations. *Quantitative Economics*, *14*(2), 609–650.

Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2016). A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica*, *84*(2), 571–626.

Kohlhas, A. N., & Walther, A. (2021). Asymmetric Attention. *American Economic Review*, *111*(9), 2879–2925.

Kučinskas, S., & Peters, F. S. (2022). Measuring Under- and Overreaction in Expectation Formation. *The Review of Economics and Statistics*, 1–45.

Maćkowiak, B., & Wiederholt, M. (2009). Optimal Sticky Prices under Rational Inattention. *American Economic Review*, *99*(3), 769–803.

Malmendier, U., & Nagel, S. (2016). Learning from inflation experiences. *Quarterly Journal of Economics*, *131*(1).

Malmendier, U., & Wachter, J. A. (2021). Memory of Past Experiences and Economic Decisions. *SSRN Electronic Journal*.

Mankiw, N. G., & Reis, R. (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, J., Wu, J., & Young, E. R. (2022). Multivariate Rational Inattention. *Econometrica*, *90*(2), 907–945.

Nagel, S., & Xu, Z. (2022). Asset Pricing with Fading Memory. *The Review of Financial Studies*, *35*(5), 2190–2245.

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

Orphanides, A., & Williams, J. C. (2006). Monetary Policy with Imperfect Knowledge. *Journal of the European Economic Association*, *4*(2-3), 366–375.

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford university press.

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

Pfäuti, O. (2025). The Inflation Attention Threshold and Inflation Surges.

Purves, D., Brannon, E. M., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., & Woldorff, M. G. (2013). *Principles of cognitive neuroscience* (2nd). Sinauer Associates.

Salle, I., Gorodnichenko, Y., & Coibion, O. (2023). Lifetime memories of inflation: Evidence from surveys and the lab. *National Bureau of Economic Research, Working Paper Series*(31996).

Shiller, R. J., & Thompson, A. K. (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*.

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

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

Survey of Professional Forecasters. (1968–2024). Historical Data Files for the Survey of Professional Forecasters [https://www.philadelphiafed.org/research-and-data/real-time-center/survey-of-professional-forecasters, Accessed on January 31, 2025].

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

Wachter, J. A., & Kahana, M. J. (2023). Associative Learning and Representativeness. *SSRN Electronic Journal*.

Wachter, J. A., & Kahana, M. J. (2024). A Retrieved-Context Theory of Financial Decisions. *The Quarterly Journal of Economics*, *139*(2), 1095–1147.

Wang, C. (2021). Under- and Overreaction in Yield Curve Expectations. *SSRN Electronic Journal*.

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

Woodford, M. (2003). Imperfect Common Knowledge and the Effects of Monetary Policy. In M. Woodford, P. Aghion, R. Frydman, & J. Stiglitz (Eds.), *Knowledge, information, and expectations in modern macroeconomics: In honor of edmund s. phelps* (pp. 25–58). Princeton University Press.

Woodford, M. (2020). Modeling Imprecision in Perception, Valuation, and Choice. *Annual Review of Economics*, *12*(1), 579–601.

Zarnowitz, V. (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 Optimal Information Selection

The state vector x_t evolves according to the following law of motion:

$$x_t = d + A\, x_{t-1} + B\, \epsilon_t, \quad \epsilon_t \sim \mathcal{N}\left(0,\, \Sigma_\epsilon\right)$$

where d, A, B, and Σ_ϵ are constant parameters known to the decision-maker(DM). The DM's default prior over the state is given by:

$$x_t \sim \mathcal{N}\left(\mu_x,\, \Sigma_x\right) \tag{A.1}$$

The DM's subjective probability measure reflects this default prior, along with beliefs about the cognitive states $m_{i,t}$ and $n_{i,t}$. This perceived distribution governs how the DM forms expectations and may differ from the actual distribution of outcomes.

In what follows, I characterize the optimal cognitive process—defined by the sequence $\{K_t, \Sigma_{ut}, \Lambda_t, \Sigma_{\omega t}\}_{t=0}^{\infty}$—that minimizes the loss function (1.9), subject to the constraints imposed by the information environment (1.2)- (1.3) and (1.5)- (1.6).

A.1 Optimal Information Selection

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 $E\left[x_t \mid M_{i,t}\right]$ with noise while the optimal $n_{i,t}$ records $E\left[x_t \mid N_t\right]$ 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

$$\begin{pmatrix} \vec{m}_{i,t} \\ \tilde{m}_{i,t} \end{pmatrix} = \begin{pmatrix} \Lambda_{at} & \Lambda_{b,t} \\ \Lambda_{ct} & \Lambda_{dt} \end{pmatrix} \begin{pmatrix} \vec{M}_{i,t} \\ E\left[x_t \mid M_{i,t}\right] \end{pmatrix} + \begin{pmatrix} \vec{\omega}_{i,t} \\ \tilde{\omega}_{i,t} \end{pmatrix} \tag{A.2}$$

where the elements of $\vec{M}_{i,t}$ are orthogonal to $E\left[x_t \mid M_{i,t}\right]$ and $\vec{M}_{i,t}$ and $E\left[x_t \mid M_{i,t}\right]$ span the same vector space as $M_{i,t}$. The terms $\vec{\omega}_{i,t}$ and $\tilde{\omega}_{i,t}$ are Gaussian innovations uncorrelated with $M_{i,t}$. I impose the normalization assumption that the vector $E\left[x_t \mid M_{i,t}\right] - \tilde{m}_{i,t}$ is uncorrelated with all the elements in $m_{i,t}$. Two requirements summarize this relationship.

$$Cov\left[E\left[x_t|\, M_{i,t}\right]-\tilde{m}_{i,t},\vec{m}_{i,t}\right]=\vec{0} \tag{A.3a}$$

$$Cov\left[E\left[x_t|\, M_{i,t}\right]-\tilde{m}_{i,t},\tilde{m}_{i,t}\right]=0 \tag{A.3b}$$

The second requirement implies that

$$\begin{aligned} & Cov\left[E\left[x_t|\, M_{i,t}\right],\tilde{m}_{i,t}\right]=Var\left[\tilde{m}_{i,t}\right] \\ \Leftrightarrow\quad & Var\left[E\left[x_t|\, M_{i,t}\right]\right]\,\Lambda'_{dt}=\Lambda_{dt}\,Var\left[E\left[x_t|\, M_{i,t}\right]\right]\,\Lambda'_{dt}+Var\left[\Lambda_{ct}\,\vec{M}_{i,t}+\tilde{\omega}_{i,t}\right] \\ \Leftrightarrow\quad & Var\left[\Lambda_{ct}\,\vec{M}_{i,t}+\tilde{\omega}_{i,t}\right]=\left(1-\Lambda_{dt}\right)Var\left[E\left[x_t|\, M_{i,t}\right]\right]\,\Lambda'_{dt} \end{aligned} \tag{A.4}$$

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

$$\begin{pmatrix}\vec{n}_{i,t}\\ \tilde{n}_{i,t}\end{pmatrix}=\begin{pmatrix}K_{at} & K_{bt}\\ K_{ct} & K_{dt}\end{pmatrix}\begin{pmatrix}\vec{N}_t\\ E\left[x_t|\, N_t\right]\end{pmatrix}+\begin{pmatrix}\vec{u}_{i,t}\\ \tilde{u}_{i,t}\end{pmatrix} \tag{A.5}$$

where $\vec{N}_t$ is the components in N_t that are orthogonal to $E\left[x_t|\, N_t\right]$, thus $E\left[x_t|\, N_t\right]$ and that $\vec{N}_t$ and $E\left[x_t|\, N_t\right]$ span the same vector space as N_t. Both $\vec{u}_{i,t}$ and $\tilde{u}_{i,t}$ are Gaussian innovations uncorrelated with N_t. As before, I impose the normalization assumption that the vector $E\left[x_t|\, N_t\right]-\tilde{n}_{i,t}$ is uncorrelated with all the elements in $n_{i,t}$, conditional on $m_{i,t}$. That is, the two requirements are

$$Cov\left[E\left[x_t|\, N_t\right]-\tilde{n}_{i,t},\vec{n}_{i,t}|\, m_{i,t}\right]=\vec{0} \tag{A.6a}$$

$$Cov\left[E\left[x_t|\, N_t\right]-\tilde{n}_{i,t},\tilde{n}_{i,t}|\, m_{i,t}\right]=0 \tag{A.6b}$$

We can see that (A.6b) implies

$$\begin{aligned} & Cov\left[x_t,K_{dt}\,E\left[x_t|\, N_t\right]|\, m_{i,t}\right]=Var\left[K_{ct}\,\vec{N}_t+K_{dt}\,E\left[x_t|\, N_t\right]+\tilde{u}_{i,t}\middle|\, m_{i,t}\right] \\ \Leftrightarrow\quad & Var\left[K_{ct}\,\vec{N}_t+\tilde{u}_{i,t}\middle|\, m_{i,t}\right]=Cov\left[x_t,K_{dt}\,E\left[x_t|\, N_t\right]|\, m_{i,t}\right] \\ & \qquad -K_{dt}\,Var\left[E\left[x_t|\, N_t\right]|\, m_{i,t}\right]\,K'_{dt} \end{aligned} \tag{A.7}$$

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

We observe from (A.2) 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.3a). We can furthermore see that Λ_{dt} uniquely determines the prior uncertainty $E\left[x_t \middle| m_{i,t}\right]$.

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

Likewise, we also observe from the proposed partition (A.5) 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.6a). Furthermore, we can see that K_{dt} uniquely determines the posterior uncertainty.

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

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

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

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

As discussed in Appendix C.2 of Azeredo da Silveira et al. (2024), the lower bound is equal to $I\left(\tilde{m}_{i,t}; E\left[x_t \middle| M_{i,t}\right]\right)$, which again is achieved when $\Lambda_{at} = \Lambda_{b,t} = \Lambda_{ct} = O$. Likewise,

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

whose lower bound of this mutual information is equal to $I\left(\tilde{n}_{i,t}; E\left[x_t \middle| N_t\right]\right)$. This lower bound is achieved when $K_{at} = K_{bt} = K_{ct} = 0$.

Step 4: Optimal Information Selection

Based on Step 1-3, we can derive the mental representation under optimal information selection. The internal data $M_{i,t}$ is represented as a function of a matrix loading $\tilde{\Lambda}_t$.

$$m_{i,t} = \tilde{\Lambda}_t\, E\left[x_t \middle| M_{i,t}\right] + \omega_{i,t}, \quad \omega_{i,t} \sim \mathcal{N}\left(0,\, \tilde{\Sigma}_{\omega,t}\right) \tag{A.8a}$$

$$\tilde{\Sigma}_{\omega,t} = \left(I - \tilde{\Lambda}_t\right) Var\left[E\left[x_t \middle| M_{i,t}\right]\right]\, \tilde{\Lambda}_t' \tag{A.8b}$$

2 The feasible set of $\tilde{\Lambda}_t$ is defined as the collection of Λ_t under which the resulting $\tilde{\Sigma}_{\omega,t}$ is a proper variance-covariance matrix (that is, symmetric and p.s.d.).

Likewise, the external data N_t is represented as a function of a matrix loading $\tilde{K}_t$.

$$n_{i,t} = \tilde{K}_t\, E\left[x_t \middle| N_t\right] + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}\left(0,\, \tilde{\Sigma}_{u,t}\right) \tag{A.9a}$$

$$\tilde{\Sigma}_{u,t} = Cov\left[x_t, E\left[x_t \middle| N_t\right] \middle| m_{i,t}\right] \tilde{K}_t' - \tilde{K}_t\, Var\left[E\left[x_t \middle| N_t\right] \middle| m_{i,t}\right]\, \tilde{K}_t' \tag{A.9b}$$

The feasible set of $\tilde{K}_t$ is defined as a collection of $\tilde{K}_t$ that yields the resulting $\tilde{\Sigma}_{u,t}$ to be a proper variance-covariance matrix (that is, symmetric and p.s.d.).

A.2 Beliefs Evolution under Optimal Information Selection

I derive the evolution of the DM's beliefs under the optimal information selection rule described in Section A.1, assuming the news vector is characterized as (1.14).

Using the notations for the default prior (A.1) and the time-t prior (1.7), I express the representations (A.8a)-(A.8b).

$$m_{i,t} = \tilde{\Lambda}_t\, x^m_{i,t|t} + \omega_{i,t}, \quad \omega_{i,t} \sim \mathcal{N}\left(0,\, \tilde{\Sigma}_{\omega,t}\right) \tag{A.10a}$$

$$\tilde{\Sigma}_{\omega,t} = (I - \tilde{\Lambda}_t)\,(\Sigma_x - \Sigma_{t|t-1})\,\tilde{\Lambda}_t' \tag{A.10b}$$

Using the notation for the time-t posterior belief in (1.8), I rewrite the representations in (A.9a)-(A.9b).

$$n_{i,t} = \tilde{K}_t\,(x_t + \nu_t) + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}\left(0,\, \tilde{\Sigma}_{u,t}\right) \tag{A.11a}$$

$$\tilde{\Sigma}_{u,t} = \Sigma^m_{t|t}\,\tilde{K}_t' - \tilde{K}_t\,(\Sigma^m_{t|t} + \Sigma_{\nu,t})\,\tilde{K}_t' \tag{A.11b}$$

Based on these expressions, the first moment of the time-t prior beliefs is derived as below:

$$\begin{aligned} x^m_{i,t|t} &= E\left[x_t\right] + Cov\left[x_t, m_{i,t}\right]\, Var\left[m_{i,t}\right]^{-1}\left(m_{i,t} - E\left[m_{i,t}\right]\right) \\ &= E\left[x_t\right] + m_{i,t} - E\left[m_{i,t}\right] \\ &= (I - \tilde{\Lambda}_t)\,\mu_x + \tilde{\Lambda}_t\, x^m_{i,t|t} + \tilde{\omega}_{i,t} \end{aligned} \tag{A.12}$$

where the second equation holds because $Cov\left[x_t, m_{i,t}\right] = Var\left[m_{i,t}\right] = Var\left[x^m_{i,t|t}\right]\, \tilde{\Lambda}_t'$ after substituting (A.10a)-(A.10b). The second moment of the prior beliefs is derived using this relationship.

$$\begin{aligned} \Sigma^m_{t|t} &= Var\left[x_t\right] - Cov\left[x_t, m_{i,t}\right]\, Var\left[m_{i,t}\right]^{-1}\, Cov\left[m_{i,t}, x_t\right] \\ &= Var\left[x_t\right] - Cov\left[m_{i,t}, x_t\right] \\ &= Var\left[x_t\right] - \tilde{\Lambda}_t\, Cov\left[\Sigma_{t|t-1}, x_t\right] \\ &= \Sigma_{t|t-1} + (I - \tilde{\Lambda}_t)\,(\Sigma_x - \Sigma_{t|t-1}) \end{aligned} \tag{A.13}$$

Similarly, the first moment of the time-t posterior beliefs is derived as:

$$\begin{aligned} x_{i,t|t} &= E\left[x_t \middle| m_{i,t}\right] + Cov\left[x_t, n_{i,t} \middle| m_{i,t}\right] Var\left[n_{i,t} \middle| m_{i,t}\right]^{-1}\left(n_{i,t} - E\left[n_{i,t} \middle| m_{i,t}\right]\right) \\ &= E\left[x_t \middle| m_{i,t}\right] + n_{i,t} - E\left[n_{i,t} \middle| m_{i,t}\right] \\ &= (I - \tilde{K}_t)\, x^m_{i,t|t} + \tilde{K}_t\,(x_t + \nu_t) + \tilde{u}_{i,t} \end{aligned} \tag{A.14}$$

where the second equation holds because $Cov\left[x_t, n_{i,t}\middle| m_{i,t}\right] = Var\left[n_{i,t}\middle| m_{i,t}\right] = \Sigma^m_{t|t}\tilde{K}'_t$ from (A.11a)-(A.11b). Using this relationship, I derive the second moment.

$$
\begin{aligned}
\Sigma_{t|t} &= Var\left[x_t\middle| m_{i,t}\right] + Cov\left[x_t, n_{i,t}\middle| m_{i,t}\right] Var\left[n_{i,t}\middle| m_{i,t}\right]^{-1} Cov\left[n_{i,t}, x_t\middle| m_{i,t}\right] \\
&= Var\left[x_t\middle| m_{i,t}\right] - Cov\left[n_{i,t}, x_t\middle| m_{i,t}\right] \\
&= Var\left[x_t\middle| m_{i,t}\right] - \tilde{K}_t Cov\left[x_t, x_t\middle| m_{i,t}\right] \\
&= (I - \tilde{K}_t)\,\Sigma^m_{t|t} \qquad \text{(A.15)}
\end{aligned}
$$

B Special Case: A Single State Variable

This section examines the optimal cognitive process in the special case where the only state variable relevant for forecasting is z_t. This simplifies the setting from the previous section, where the state vector x_t was allowed to take a more general form. In this case, the news vector is given by equation (2.4).

B.1 Optimal representation of noisy news

The optimal $n_{i,t}$ takes the form

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

for some positive scalar κ_{zt} to be determined. The idiosyncratic noise $\tilde{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}(0,\, \sigma^2_{ut})$.

The values of κ_{zt} and σ^2_{ut} that satisfy the normalization assumption (A.3b) and the information constraint (1.3) are:

$$\kappa_{zt} = \frac{\Sigma^m_{z,t|t}}{\Sigma^m_{z,t|t} + \sigma^2_\nu + \frac{\phi_n}{1-\phi_n}\left(\sigma^2_z + \sigma^2_\nu\right)} \tag{B.16}$$

$$\sigma^2_{ut} = \kappa^2_{zt}\left(\frac{\phi_n}{1-\phi_n}\right)\left(\sigma^2_z + \sigma^2_\nu\right) \tag{B.17}$$

where $\sigma^2_z \equiv Var\left[z_t\right]$.

B.2 Optimal representation of noisy memory

Likewise, the optimal $m_{i,t}$ is given by

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

for some positive scalar λ_t. The idiosyncratic noise term $\tilde{\omega}_{i,t}$ follows a Gaussian distribution $\mathcal{N}(0,\, \sigma^2_{\omega t})$.

The normalization assumption (A.3b) and the information constraint (1.6) pin down the values of λ_t and $\sigma^2_{\omega t}$ as below.

$$\lambda_t = 1 - \phi_m \tag{B.18}$$

$$\sigma^2_{\omega t} = \phi_m\left(1-\phi_m\right) Var\left[z_{i,t|t-1}\right] \tag{B.19}$$

The simplicity of this solution arises because memory resources need not be allocated across multiple variables.

C Derivation of Optimal Information Transformation

This section derives the optimal information transformation when the full state vector x_t is subject to cognitive processing. I characterize the structure of the attention and memory systems that jointly govern belief formation, with particular focus on how memory resources are optimally allocated across the components of x_t.

C.1 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

$$N_t = c'\, x_t + \bar{\nu}_t, \quad \bar{\nu}_t \sim \mathcal{N}\left(0,\, \sigma_\nu^2\right) \tag{C.20}$$

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

$$\tilde{n}_{i,t} = \kappa_t\, \bar{n}_{i,t} \tag{C.21}$$

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

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

for some positive scalar κ_{ct} that remains to be specified, and the idiosyncratic noise $\bar{u}_{i,t}$ follows a Gaussian distribution $\mathcal{N}\left(0,\, \sigma_{ut}^2\right)$. In summary, the noisy representation of external data is described as (A.11a) 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.6b) 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

$$e_i'\, \kappa_t = \frac{\kappa_{ct}\left(e_i'\, \Sigma_{t|t}^m\, c\right)}{\kappa_{ct}^2 (c' \Sigma_{t|t}^m\, c + \sigma_\nu^2) + \sigma_{ut}^2} \tag{C.22}$$

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

$$\sigma_{ut}^2 = \kappa_{ct} (e_1' \Sigma_{t|t}^m c) - \kappa_{ct}^2 (c' \Sigma_{t|t}^m c + \sigma_\nu^2). \tag{C.23}$$

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 (C.23) into the (C.22) results in the expression for the column vector κ_t as

$$\kappa_t = \frac{1}{e_1' \, \Sigma_{t|t}^m \, c} \, \Sigma_{t|t}^m \, c. \tag{C.24}$$

Using the information constraint, we can derive that

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

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

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

Substituting (C.23) yields that

$$\kappa_{ct} \leq \frac{e_1' \Sigma_{t|t}^m c}{c' \Sigma_{t|t}^m c + \sigma_\nu^2 + \frac{\phi_n}{1-\phi_n} \left(c' \Sigma_x c + \sigma_\nu^2\right)} \tag{C.25}$$

Thus the optimal value of κ_{ct} corresponds to the upper bound in (C.25). Then, the resulting σ_{ut}^2 is

$$\sigma_{ut}^2 = \kappa_{ct}^2 \left(\frac{\phi_n}{1 - \phi_n}\right) \left(c' \Sigma_x c + \sigma_\nu^2\right) \tag{C.26}$$

C.2 Specification of $\tilde{\Lambda}_t$

The mutual-information capacity (1.6) 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 $\Sigma^m_{t|t}$ is a symmetric and positive semidefinite matrix. Second, the diagonal elements of $\Sigma^m_{t|t}$ are bigger than those of $\Sigma_{t|t-1}$ and smaller than those of Σ_x. That is, under any feasible $\tilde{\Lambda}_t$, both $\Sigma^m_{t|t} - \Sigma_{t|t-1}$ and $\Sigma_x - \Sigma^m_{t|t}$ are proper variance-covariance matrices (symmetric and positive semidefinite).

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

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

We could confirm that the same accuracy constraint (1.6) applies.

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

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

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

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

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

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

The Lagrangian Problem and the First Order Conditions

The Lagrangian problem is as follows.

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

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

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

$$-\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}\Gamma_1\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}} + \Upsilon_1 - \Upsilon_2 + \mu\,\det\left(W_t\right)W_t^{-1} = 0 \tag{C.27a}$$

$$\Gamma_1 - c\,\Gamma_2\,c' - \Gamma_3 + c\left(\Omega^m_{t|t}\right)^{-1} c'\,\Sigma^m_{t|t}\,\Gamma_3 + \Gamma_3\,\Sigma^m_{t|t}\,c\left(\Omega^m_{t|t}\right)^{-1} c' = 0 \tag{C.27b}$$

$$\Gamma_2 - \left(\Omega^m_{t|t}\right)^{-1}\,c'\,\Sigma^m_{t|t}\,\Gamma_3\,\Sigma^m_{t|t}\,c\,\left(\Omega^m_{t|t}\right)^{-1} = 0 \tag{C.27c}$$

$$-\tilde{Q} + \Gamma_3 = 0 \tag{C.27d}$$

and the slackness conditions are

$$\Upsilon_1 W_t = 0,\ \Upsilon_1 \succeq 0,\ W_t \succeq 0 \tag{C.28a}$$

$$\Upsilon_2\left(I - W_t\right) = 0,\ \Upsilon_2 \succeq 0,\ \left(I - W_t\right) \succeq 0 \tag{C.28b}$$

and

$$\mu\,\left(\det\left(W_t\right) - \phi_m\right) = 0,\ \mu \geq 0, \det\left(W_t\right) = \phi_m \tag{C.29}$$

We can first rearrange (C.27b)-(C.27d). Note that $\Gamma_3 = Q$ (as implied by (C.27d)) and using the notation $\tilde{K}_t \equiv \Sigma^m_{t|t}\,c\,(\Omega^m_{t|t})^{-1}c'$, we can express (C.27b) as

$$\Gamma_1 - c\,\Gamma_2\,c' - \tilde{Q} + \tilde{K}'_t\tilde{Q} + \tilde{Q}\,\tilde{K}_t = 0$$

and (C.27c) as

$$c\,\Gamma_2\,c' - \tilde{K}'_t\,\tilde{Q}\,\tilde{K}_t = 0$$

which together result in

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

Next, I'd like to solve for W_t that characterizes the optimal memory system. First, multiplying (C.27a) by $W_t\ (I - W_t)$ on the left yields

$$-\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}\Gamma_1\left(\Sigma_x - \Sigma_{t|t-1}\right)^{\frac{1}{2}}\ W_t\ \left(I - W_t\right) + \mu\,\phi_m\ \left(I - W_t\right) = 0 \tag{C.30}$$

after applying the slackness conditions (from which $(\Upsilon_1 - \Upsilon_2) W_t \, (I - W_t) = 0$). 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

$$U \, \left(\mu \, \phi_m \, I - G \, (I - D)\right) D \, U' = 0 \tag{C.31}$$

Note that D should satisfy $D \succeq 0$, $I - D \succeq 0$, 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 $\Sigma_{t|t}^m$ is derived as

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

where $Var\left[x_{i,t|t-1}\right] = \Sigma_x - \Sigma_{t|t-1}$ captures the maximum possible increase in the uncertainty due to forgetting the previous 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 $Var\left[x_{i,t|t-1}\right]$).

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

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

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

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

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

D Forecast Dynamics and Bias in the Univariate Case

This section analyzes the stationary relationship between beliefs and the underlying state in the univariate case where the only state variable is z_t. I compute the dynamics of forecasts and derive the resulting forecast bias implied by the optimal cognitive process. The simplicity of this setting allows for closed-form expressions that highlight key mechanisms.

D.1 Derivations of the regression coefficients β_I and β_C

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

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

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

$$z_{t|t} = (1-\lambda)(1-\kappa_z)\,\mu + \lambda\,(1-\kappa_z)\,z_{t|t-1} + \kappa_z\,z_t + \kappa_z\,\tilde{\nu}_t \tag{D.32}$$

I define b as the weight on unconditional prior belief.

$$b \equiv (1-\lambda)(1-\kappa_z) \tag{D.33}$$

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

$$\beta_I = -\frac{b}{2\,b + \rho^{-2} - 1} \tag{D.34}$$

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

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

Therefore, it is straightforward to see that

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

We can also show that

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

And finally,

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

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

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

Then,

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

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

$$\beta_C = \frac{1 - \kappa_z}{\kappa_z}\left(1 + (1 - \lambda)\,\frac{\lambda\left(1 - \kappa_z\right)\rho^2}{1 - \lambda\left(1 - \kappa_z\right)\rho^2}\right) \tag{D.35}$$

D.2 Steady-state Uncertainty

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

$$\Sigma_{-1} = \rho^2\,\Sigma + \sigma_\epsilon^2 \tag{D.36a}$$

$$\Sigma^m = (1 - \lambda)\,\sigma_z^2 + \lambda\,\Sigma_{-1} \tag{D.36b}$$

$$(\Sigma)^{-1} = (\Sigma^m)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} \tag{D.36c}$$

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

$$\kappa_z = \frac{\Sigma^m}{\Sigma^m + \tilde{\sigma}_u^2} \tag{D.37}$$

$$b = (1 - \lambda)\,\frac{\tilde{\sigma}_u^2}{\Sigma^m + \tilde{\sigma}_u^2} \tag{D.38}$$

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

D.3 Comparative Statics

Comparative statics for the posterior uncertainty. Equations (D.36) implicitly impose the following relation.

$$\begin{aligned} F\left(\Sigma; \tilde{\sigma}_u^2, \lambda\right) &= (\Sigma^m)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - (\Sigma)^{-1} \\ &= \left((1 - \lambda)\,\sigma_z^2 + \lambda\,\left(\rho^2\,\Sigma + \sigma_\epsilon^2\right)\right)^{-1} + \left(\tilde{\sigma}_u^2\right)^{-1} - (\Sigma)^{-1} = 0 \end{aligned} \tag{D.39}$$

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

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

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

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

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

$$\frac{\partial \Sigma}{\partial \lambda} = - \left(\left(\frac{\Sigma^m}{\Sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \left(\sigma_z^2 - \Sigma_{-1} \right)$$

$$= - \left(\left(\frac{\Sigma^m}{\sigma} \right)^2 - \lambda\, \rho^2 \right)^{-1} \frac{\Sigma^m}{1-\lambda} \left(1 - \frac{\Sigma_{-1}}{\Sigma^m} \right) < 0$$

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:

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

$$= - \frac{\Sigma^m}{1-\lambda} \left(1 - \frac{\Sigma_{-1}}{\Sigma^m} \right) \left\{ 1 + \frac{\lambda\, \rho^2}{\left(\frac{\Sigma^m}{\Sigma}\right)^2 - \lambda \rho^2} \right\}$$

$$= - \frac{\Sigma^m}{1-\lambda} \frac{1 - \frac{\Sigma_{-1}}{\Sigma^m}}{1 - \lambda \rho^2 \left(\frac{\Sigma}{\Sigma^m}\right)^2} < 0$$

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:

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

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

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:

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

$$= - \frac{\tilde{\sigma}_u^2}{\Sigma_m + \tilde{\sigma}_u^2} \left(1 - \frac{\Sigma^m}{\Sigma_m + \tilde{\sigma}_u^2} \frac{1 - \frac{\Sigma_{-1}}{\Sigma^m}}{1 - \lambda \rho^2 \left(\frac{\Sigma}{\Sigma^m}\right)^2} \right) < 0$$

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

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

Finally, the derivative of κ_z with respect to λ:

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

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 (D.34) 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

$$\frac{\partial \beta_I}{\partial \phi_m} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \phi_m} = -\frac{\partial \beta_I}{\partial b}\frac{\partial \beta_I}{\partial \lambda} < 0 \tag{D.40a}$$

$$\frac{\partial \beta_I}{\partial \phi_n} = \frac{\partial \beta_I}{\partial b}\frac{\partial b}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} < 0 \tag{D.40b}$$

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 (D.35), 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

$$\frac{\partial \beta_C}{\partial \phi_n} = \frac{\partial \beta_C}{\partial \kappa_z}\frac{\partial \kappa_z}{\partial \tilde{\sigma}_u^2}\frac{\partial \tilde{\sigma}_u^2}{\partial \phi_n} > 0 \tag{D.41}$$

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 $\tilde{\sigma}_u^2 = \frac{\phi_n}{1-\phi_n}\sigma_z^2$.

E Long-run Uncertainty: Inspecting the Mechanism

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 over μ

$$\mu \sim \mathcal{N}(\bar{\mu}, \Omega). \tag{E.42}$$

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

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

where 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.[25]

However, even with infinitely long data series, forecasters' uncertainty about μ may not vanish completely if memory frictions limit their access to past information. As discussed by Azeredo da Silveira et al. (2024), noisy memory prevents DM from accurately retrieving internal information, keeping $Var\left[\mu | m_{i,t}, n_{i,t}\right]$ bounded away from zero even after prolonged learning. A similar intuition is developed in Nagel and Xu (2022).

Why does persistent uncertainty about the long-run mean matter? It matters because, despite knowing the mean is constant, the DM will continually update beliefs about μ in response to new data. Consequently, a high realization of z_t partially shifts beliefs toward a higher-than-expected μ, leading the forecaster to predict persistently elevated future values of z_t.

Error-revision regression. Perpetual uncertainty about the long-run mean also implies that the regression coefficients in the forecast error-revision test (2.11) and (2.10) will vary across forecast horizons. To illustrate, consider the regression coefficient applied to forecasts for μ. Defining the individual 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 have:

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

[25] 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).

The derivation is straightforward. Specifically, one observes that

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

A similar derivation holds for β_I.[26]

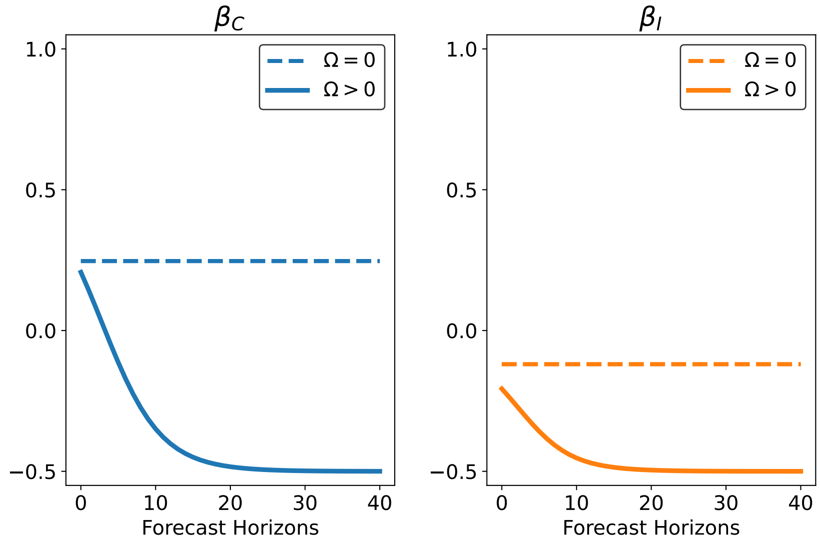
Figure 10 illustrates the model's predictions for β_C and β_I for varying forecast horizons. I fix the degree of attention friction and memory friction at the levels used in Figure 1, calibrated to match the targeted β_C and β_I. All other parameters follow the configuration in the left panel of Figure 11. The figure shows that both coefficients become increasingly negative as the forecast horizon lengthens. As previously derived, for sufficiently long horizons, β_C^μ and β_I^μ approach $-\frac{1}{2}$.

Impulse response function. Figure 11 illustrates the impulse response to a transitory shock in z_t, using the same data-generating process as in Figure 1, with cognitive parameters set to $\phi_n = 0.2$, $\phi_m = 0.2$, and $\Omega = 0.5 \times Var\left[z_t\right]$. The left panel shows responses to an innovation in z_t, with the black dashed line representing the actual path of z_t. The blue line shows how forecasts for z_t evolve following the innovation, highlighting continued sluggishness in belief updates due to attention frictions. The orange line shows forecasts for μ: the DM partly attributes high values of z_t to a higher long-run mean.

The right panel of Figure 11 displays the response of four-quarter-ahead forecasts for varying degrees of Ω. I realign the forecasts relative to the realized z_{t+4} (black dashed line) to clearly illustrate forecast undershooting or overshooting. Initially, forecasts undershoot across all values of Ω due to attention frictions. However, forecasts begin to overshoot after a few periods for larger values of Ω. When Ω is high, the DM excessively revises her beliefs about the long-run mean, offsetting initial undershooting. Consequently, forecast errors, defined as $z_{t+4} - z_{i,t+4|t}$, are initially positive but turn negative shortly afterward.

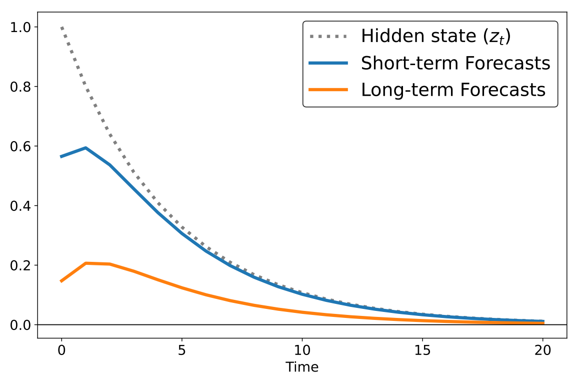
[26] Derivations for other horizons are provied in Appendix F.

Figure 10: β_C and β_I when learning about the long run

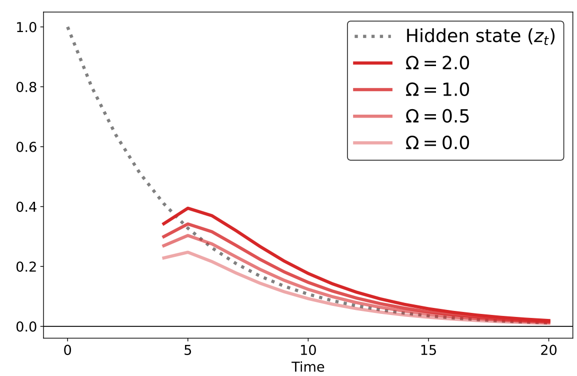


This figure shows model predictions of the two regression coefficients in (2.11) and (2.10) for different forecast horizons. The extent of cognitive noise is set as $\phi_n^* = 0.35$ and $\phi_m^* = 0.35$ (from Figure 1). Remaining parameters are from the top panel of Figure 11. The solid line assumes no uncertainty about the long-run mean($\Omega = 0$). The dashed line is when DM learns about the long run ($\Omega > 1$).

Figure 11: 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.

F Forecast Dynamics and Bias in the General Case

This section extends the analysis of forecast dynamics and bias to the general case where the full state vector x_t is subject to cognitive processing. I characterize how attention and memory jointly shape the evolution of beliefs and derive the resulting forecast bias.

Beliefs evolution. The law of motion of the posterior mean for x_t is derived as follows by combining equations (A.12) and (A.14).

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

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

$$noise_{i,t} \equiv \left(I - \tilde{K}_t\right)\tilde{\omega}_{i,t} + \tilde{u}_{i,t}. \tag{F.43}$$

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

$$x_{i,t|t} = \Delta_t\,\mu_x + \hat{A}_t\, x_{i,t-1|t-1} + \tilde{K}_t\, x_t + \tilde{\nu}_t + noise_{i,t}. \tag{F.44}$$

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

$$x_{i,t|t} = \Delta\,\mu_x + \hat{A}\, x_{i,t-1|t-1} + \tilde{K}x_t + \tilde{\nu}_t + noise_{i,t} \tag{F.45}$$

The consensus forecasts $x_{t|t}$ average the individual forecasts $x_{i,t|t-1}$ across the continuum of forecaster i in (F.45) at each time t. Thus, $x_{t|t}$ evolves according to

$$x_{t|t} = \Delta\,\mu_x + \hat{A}\, x_{t-1|t-1} + \tilde{K}x_t + \tilde{\nu}_t, \tag{F.46}$$

The deviation of individual forecasts from the consensus can be derived from subtracting (F.46) from (F.45). 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

$$x_{i,t|t} - x_{t|t} = \sum_{j=0}^{\infty} \hat{A}^j \, noise_{i,t-j} \equiv NoiseHistory_{i,t} \tag{F.47}$$

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 $Var\left[noise_{i,t}\right]$.

Violation of the law of iterated expectation. Rearranging (F.45) yields the following expression

$$x_{i,t|t} = x_{i,t|t-1} + \tilde{K}\left(x_t - x_{i,t|t-1}\right) - \Delta\left(x_{i,t|t-1} - \mu_x\right) + \tilde{\nu}_t + noise_{i,t}, \tag{F.48}$$

from which one could see that

$$E\left[x_{i,t|t} \middle| \tilde{m}_{i,t-1}, \tilde{n}_{i,t-1}\right] = (I - \Delta)\, x_{i,t|t-1} + \Delta\, \mu_x. \tag{F.49}$$

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.

F.1 Subjective Forecast Dynamics

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.

$$Var\left[x_{i,t|t}\right] = \Sigma_x - \Sigma_{t|t} \tag{F.50a}$$

$$Cov\left[x_{i,t|t}, x_t\right] = Var\left[x_{i,t|t}\right] \tag{F.50b}$$

$$\begin{aligned} Cov\left[x_{i,t|t}, x_{i,t-1|t-1}\right] &= Cov\left[\hat{A}x_{i,t-1|t-1} + \tilde{K}A\, x_t, x_{i,t-1|t-1}\right] \\ &= Cov\left[\left(\hat{A} + \tilde{K}A\right) x_{i,t-1|t-1}, x_{i,t-1|t-1}\right] \\ &= (I - \Delta)\, A\, Var\left[x_{i,t|t}\right] \end{aligned} \tag{F.50c}$$

where the equation (F.50a) 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 (F.50b) uses the "Law of Total Covariance" by focusing on the role of DM's time-t cognitive state. Finally, the auto-covariance (F.50c) uses the law of motion of the posterior mean (F.45).

The same set of statistical properties of the consensus forecast is derived below. (F.47).

$$Var\left[x_{t|t}\right] = Var\left[x_{i,t|t}\right] - Var\left[NoiseHistory_{i,t}\right] \tag{F.51a}$$
$$Cov\left[x_{t|t}, x_t\right] = Cov\left[x_{i,t|t}, x_t\right] \tag{F.51b}$$
$$Cov\left[x_{t|t}, x_{t-1|t-1}\right] = \hat{A}\, Var\left[x_{t|t}\right] + \tilde{K} A\, Cov\left[x_t, x_{t|t}\right] \tag{F.51c}$$

The derivations are based on its law of motion (F.46) 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.

$$Var\left[x_{i,t|t} - x_{t|t}\right] = Var\left[NoiseHistory_{i,t}\right] \tag{F.52a}$$
$$Cov\left[x_{i,t|t} - x_{t|t}, x_{i,t-1|t-1} - x_{t-1|t-1}\right] = \hat{A}\, Var\left[x_{i,t|t} - x_{t|t}\right] \tag{F.52b}$$

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.

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

Likewise, the variance of revisions is computed as:

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

Note that all the terms constituting forecast errors and revisions can be computed from the moments derived in the equations (F.50). The corresponding statistics for the consensus forecasts is computed in a similar way.

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.

$$\beta_C = \frac{Cov\left[y_{t+h} - F_t y_{t+h}, F_t y_{t+h} - F_{t-1} y_{t+h}\right]}{F_t y_{t+h} - F_{t-1} y_{t+h}} \tag{F.53}$$

$$= \frac{\alpha_h' Cov\left[error_t, revision_t\right] \alpha_h}{\alpha_h' Var\left[revision_t\right] \alpha_h} \tag{F.54}$$

$$\beta_I = \frac{Cov\left[y_{t+h} - F_{i,t} y_{t+h}, F_{i,t} y_{t+h} - F_{i,t-1} y_{t+h}\right]}{F_{i,t} y_{t+h} - F_{i,t-1} y_{t+h}} \tag{F.55}$$

$$= \frac{\alpha_h' Cov\left[error_{it}, revision_{it}\right] \alpha_h}{\alpha_h' Var\left[revision_{it}\right] \alpha_h} \tag{F.56}$$

$$\beta_P = \frac{Cov\left[F_{i,t} y_{t+h} - F_t y_{t+h}, F_{i,t-1} y_{t+h} - F_{t-1} y_{t+h}\right]}{F_{i,t-1} y_{t+h} - F_{t-1} y_{t+h}} \tag{F.57}$$

$$= \frac{\alpha_h' Cov\left[x_{i,t|t} - x_{t|t}, x_{i,t-1|t-1} - x_{t-1|t-1}\right] \alpha_{h+1}}{\alpha_{h+1}' Var\left[x_{i,t-1|t-1} - x_{t-1|t-1}\right] \alpha_{h+1}} \tag{F.58}$$

F.2 Parameter Uncertainty: Perceived Variation vs. Actual Variability

Section E considers uncertainty about the parameter value μ. In this setting, the DM correctly believes that the very long-run mean μ is constant but is uncertain about its exact value. Because the true data-generating process is governed by a fixed μ, the perceived forecast variation derived earlier will differ from the actual variability.

For a given value of μ, the expectations of (F.45) are derived as below.

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

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

$$E\left[x_{i,t|t} \middle| \mu\right] = cons + \underbrace{\left(I - \hat{A}\right)^{-1} \tilde{K}}_{\equiv D} E\left[x_t \middle| \mu\right] \tag{F.59}$$

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

$$E\left[error_{i,t} \middle| \mu\right] = (I - D)\, E\left[x_t \middle| \mu\right] \tag{F.60}$$

$$E\left[revision_{i,t} \middle| \mu\right] = (I - A)\, D\, E\left[x_t \middle| \mu\right] \tag{F.61}$$

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

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

from the law of total covariance. Using the correction equations (F.60) and (F.61), we can derive the covariance as

$$Cov\left[error_{i,t},revision_{i,t}\right|\mu] = Cov\left[error_{i,t},revision_{i,t}\right] \tag{F.62}$$
$$-\left(I-D\right)Var\left[E\left[x_t\right|\mu\right]\right]D'\left(I-A\right)' \tag{F.63}$$

Likewise, the actual variation in forecast revisions is derived as

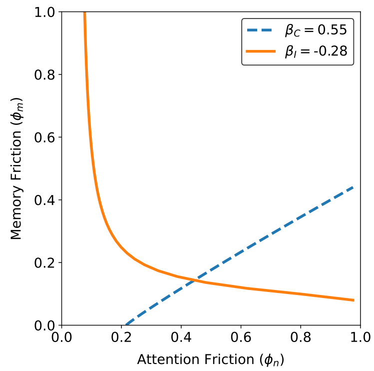
$$Var\left[revision_{i,t}\right|\mu] = Var\left[revision_{i,t}\right] - Var\left[E\left[revision_{i,t}\right|\mu\right]\right]$$
$$= Var\left[revision_{i,t}\right] - \left(I-A\right)D\,Var\left[E\left[x_t\right|\mu\right]\right]D'\left(I-A\right)' \tag{F.64}$$

The adjustments in (F.63) and (F.64) 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.

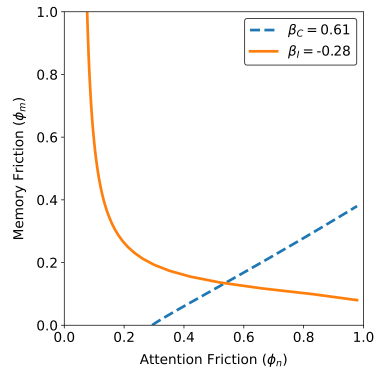
G Additional Estimation Results

This section reports additional results from Section 4. Figure 12 displays the isocurves of the estimated regression coefficients, β_C and β_I, for all macroeconomic variables considered. Figures 13-15 present the estimated impulse responses to a transitory shock across all variables.

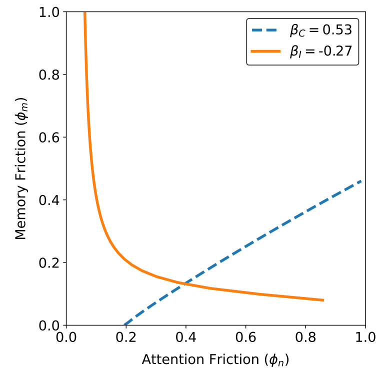
Figure 12: Estimating frictions for different macroeconomic variables



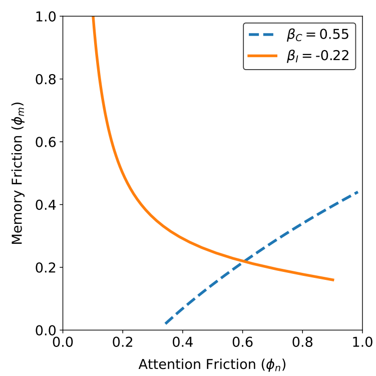
(a) CPI



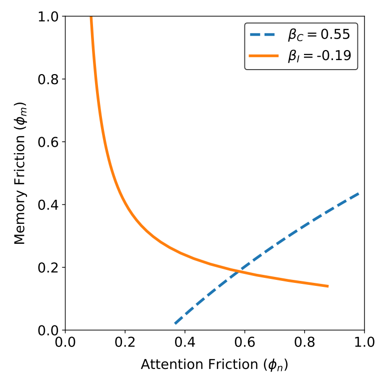
(b) PCE



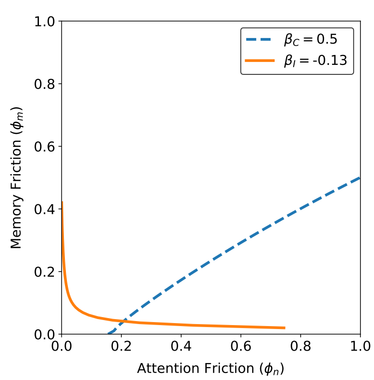
(c) PGDP



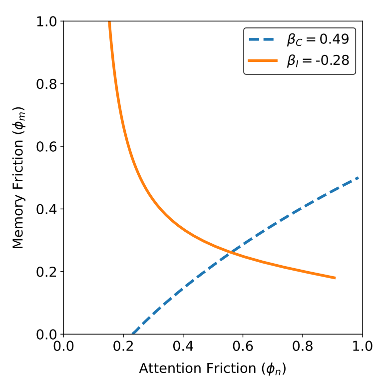
(d) NGDP



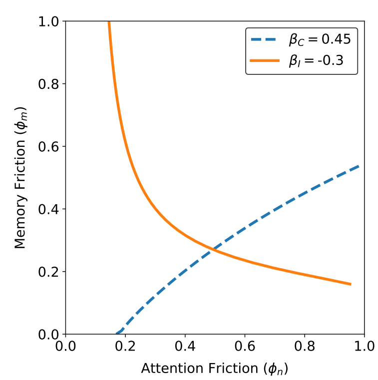
(e) RGDP



(f) TBOND

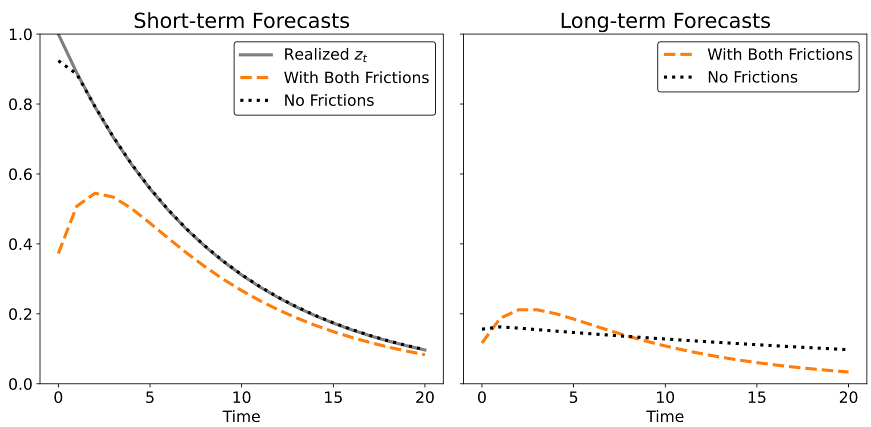


(g) BOND

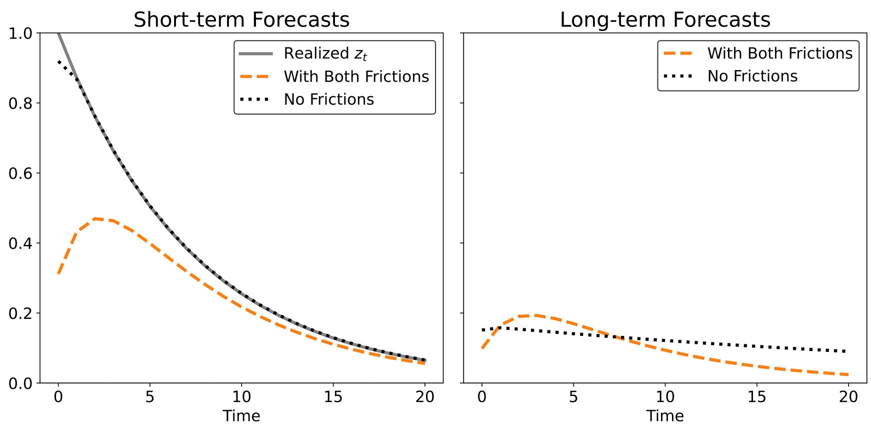


(h) BAABOND

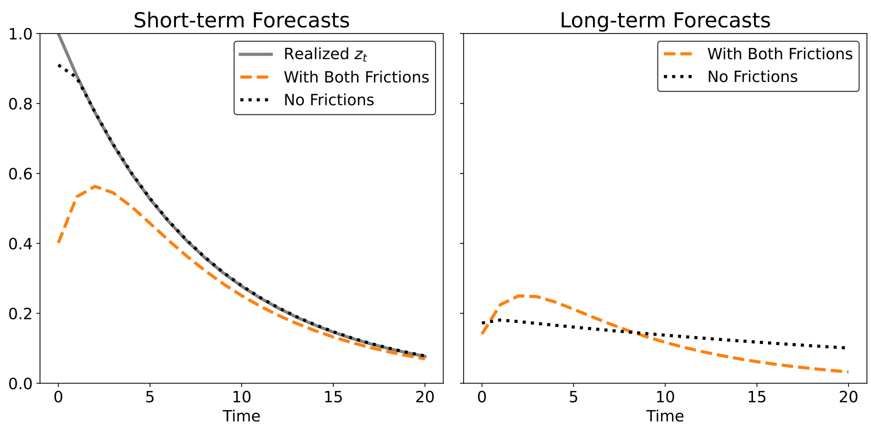
Figure 13: Estimated IRFs for inflations



(a) CPI

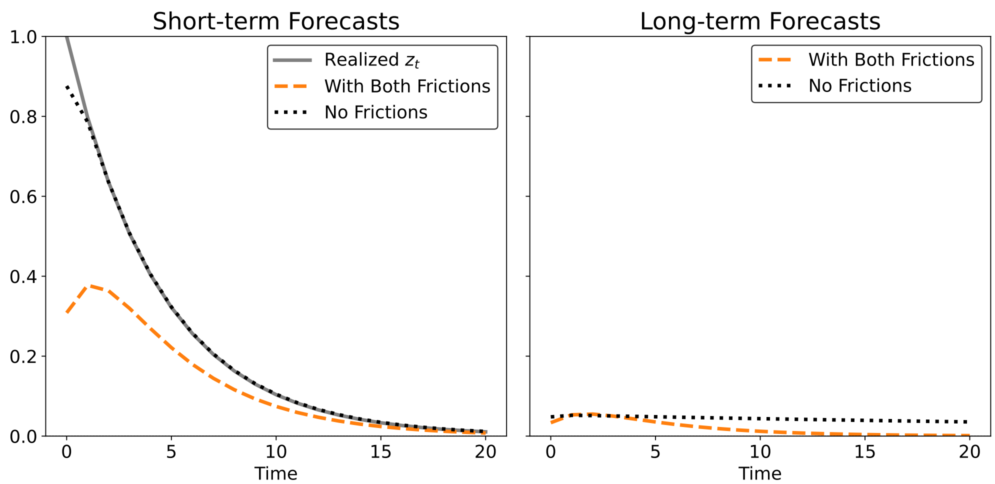


(b)PCE

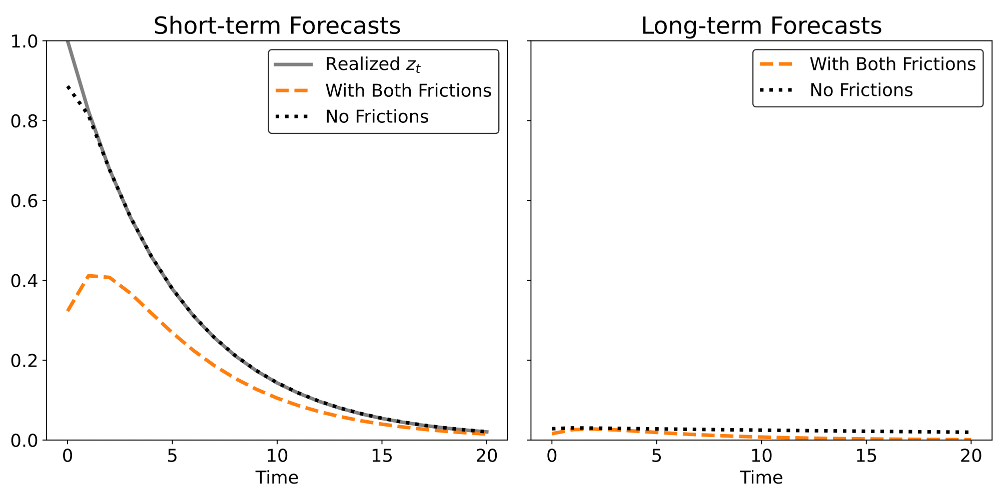


(c)PGDP

Figure 14: Estimated IRFs for output

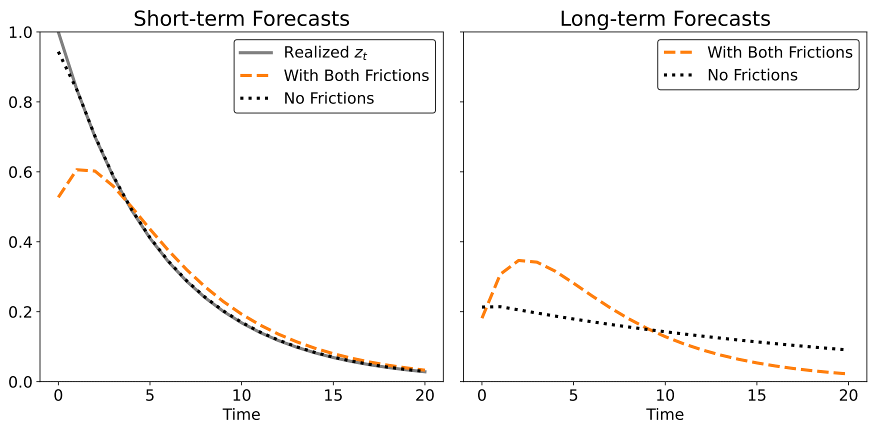


(d)NGDP

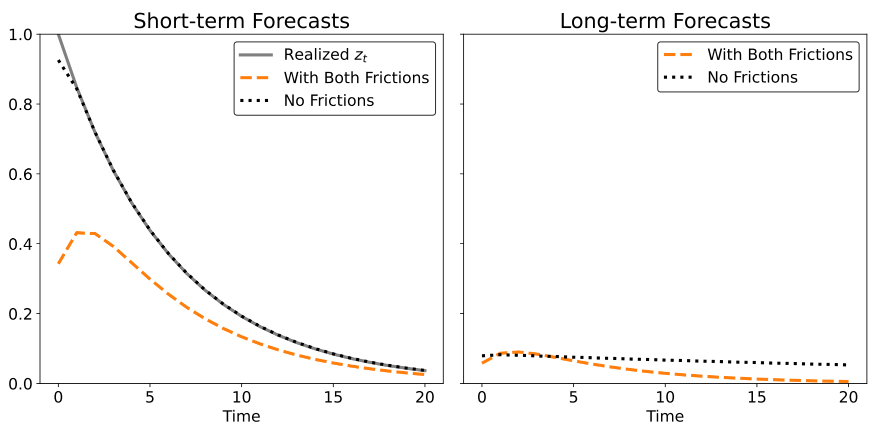


(e) RGDP

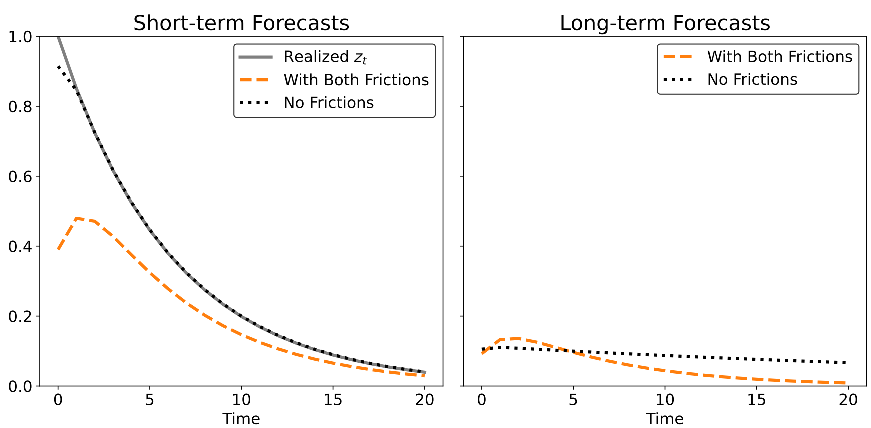
Figure 15: Estimated IRFs for interest rates



(f) TBOND



(a) BOND



(b) BAABOND

H Macroeconomic Framework

I outline a standard textbook New-Keynesian model below, following the presentation in Gali (2008, Chapter 3). I also solve for the optimal monetary policy as a linear regulator problem, taking firms' expectations as given. This setup provides a simple environment for evaluating how forecast behavior—shaped by cognitive frictions—affects policy design in the presence of trade-offs between inflation and output gap stabilization.

H.1 Household Problem

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

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

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

$$w_t - p_t = \sigma\, c_t + \varphi\, n_t \tag{H.65}$$

$$c_t = E_t\, c_{t+1} - \frac{1}{\sigma}\left(-q_t - E_t\, \pi_{t+1} + log\, \beta\right) \tag{H.66}$$

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

H.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

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

H.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,

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

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

$$\tilde{y}_t = y_t - y_t^e$$

Thus, the marginal costs are derived as

$$mc_{t+h} = \chi\, \tilde{y}_t$$

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

H.5 Optimal Monetary Policy

The cognitive model in Section 3 characterizes the evolution of average beliefs over time as follows:

$$x_{t|t} = \hat{A}\, x_{t-1|t-1} + \hat{B}\, z_t + \tilde{\nu}_t, \tag{H.67}$$

where beliefs are expressed as deviations from the steady state. To minimize notational complexity, I maintain the same notation.

The macroeconomic model in this section describes how inflation is determined by firms' beliefs about the underlying state:

$$\pi_t = d'_\pi\, x_{t|t} \tag{H.68}$$

where the row vector d'_π is given by:

$$d'_\pi = (1-\alpha) \begin{pmatrix} \frac{1}{1-\alpha\beta\rho} & \left(\frac{1}{1-\alpha\beta\rho_\mu} - \frac{1}{1-\alpha\beta\rho} \right) \end{pmatrix} \tag{H.69}$$

Additionally, the latent state z_t is determined by aggregate variables in the economy:

$$z_t = \pi_t + (1-\alpha\beta)\chi\,\tilde{y}_t + (1-\alpha\beta)\,e_t. \tag{H.70}$$

Given the model framework outlined above, the central bank influences the dynamics of the output gap and inflation.

In formulating the model economy as a linear regulator problem, the exogenous variables are defined as

$$s_t = \begin{pmatrix} x_{t-1|t-1} \\ \tilde{\nu}_t \\ e_t \end{pmatrix}$$

This state representation captures agents' beliefs from the previous period about the underlying state ($x_{t-1|t-1}$), the noise in the public information ($\tilde{\nu}_t$), and the current exogenous shock (e_t). To extract the respective components from s_t, I use selection matrices I'_x, I'_ν, and I'_e , which isolate each element as follows:

$$I'_x\,s_t = x_{t-1|t-1}; \quad I'_\nu\,s_t = \tilde{\nu}_t; \quad I'_e\,s_t = e_t.$$

Substituting equations (H.67) and (H.70) into (H.68) and rearranging, we obtain the relationship between inflation, the exogenous variables, and the output gap:

$$\pi_t = \Phi'_\pi\,s_t + \phi_\pi\,\tilde{y}_t \tag{H.71}$$

where the coefficients are given by

$$\begin{aligned} \Phi'_\pi &= \frac{d'_\pi}{1-d'_\pi\,\hat{B}}\begin{pmatrix}\hat{A} & I & (1-\alpha\beta)\hat{B}\end{pmatrix} \\ \phi_\pi &= (1-\alpha\beta)\,d'_\pi\,\hat{B} \end{aligned}$$

where I is an identity matrix with dimensions matching the size of the belief vector ($x_{t-1|t-1}$). Using this relationship and (H.70), we can express the latent variable as a function of the exogenous variables and the output gap:

$$z_t = \Phi'_z\,s_t + \phi_z\,\tilde{y}_t \tag{H.72}$$

where the coefficients are given by

$$\begin{aligned} \Phi'_z &= \Phi'_\pi + (1-\alpha\beta)\,I'_e \\ \phi_z &= \phi_\pi + (1-\alpha\beta)\chi \end{aligned}$$

Finally, using equations (H.67) and (H.72), we can expresses how the average belief $x_{t|t}$ evolves based on past beliefs, the exogenous shock e_t, the output gap $\tilde{y}_t$, and the noise term $\tilde{\nu}_t$.

$$x_{t|t} = \hat{A}\, x_{t-1|t-1} + \hat{B}\,(\Phi_z'\, s_t + \phi_z\, \tilde{y}_t) + \tilde{\nu}_t \tag{H.73}$$

$$= \left(\hat{A} + \hat{B}\,\Phi_z'\, I_x\right) x_{t-1|t-1} + \left(\hat{B}\,\Phi_z'\, I_\nu + I\right) \tilde{\nu}_t + \hat{B}\,\Phi_z'\, I_e\, e_t + \hat{B}\phi_z\, \tilde{y}_t \tag{H.74}$$

where we use the identity

$$s_t = I_x\, x_{t-1|t-1} + I_\nu\, \tilde{\nu}_t + I_e\, e_t$$

Using the law of motion of the beliefs, we can describe the evolution of the exogenous state:

$$s_{t+1} = A_s\, s_t + B_s\, \tilde{y}_t + C_s \begin{pmatrix} \tilde{\nu}_{t+1} \\ \varepsilon_{e,t+1} \end{pmatrix} \tag{H.75}$$

where the matrices A_s, B_s and C_s are given by:

$$A_s = \begin{pmatrix} \hat{A} + \hat{B}\,\Phi_z'\, I_x & \hat{B}\,\Phi_z'\, I_\nu + I & \hat{B}\,\Phi_z'\, I_e \\ 0_{2x2} & 0_{2x2} & 0 \\ 0 & 0 & \rho_e \end{pmatrix}; \quad B_s = \begin{pmatrix} \hat{B}\phi_z \\ 0_{2x2} \\ 0 \end{pmatrix}; \quad C_s = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ I & 0_{2x1} & \\ 0 & 0 & 1 \end{pmatrix} \tag{H.76}$$

The central bank's loss function is defined as the discounted sum of squared inflation and output gap deviations:

$$\min \sum_{t=0}^{\infty} \beta^t \left(\pi_t^2 + \tilde{y}_t^2\right) \tag{H.77}$$

Using equation (H.71), we can reformulate the central bank's problem as an optimal linear regulator problem, where the loss function is express in terms of the external variables s_t and the output gap $\tilde{y}_t$:

$$\max - \sum_{t=0}^{\infty} \beta^t \left\{ s_t' \Phi_\pi \Phi_\pi'\, s_t + 2 s_t' \left(\phi_\pi \Phi_\pi\right) \tilde{y}_t + \left(1 + \phi_\pi^2\right) \tilde{y}_t^2 \right\} \tag{H.78}$$

subject to the law of motion of the external variables (H.75).

The optimal output gap is given by

$$\tilde{y}_t = -F\, s_t, \tag{H.79}$$

where the matrix F is determined as

$$F = \left((1 + \phi_\pi^2) + \beta B_s' P B_s\right)^{-1} \left(\beta B_s' P A_s + \phi_\pi \Phi_\pi'\right) \tag{H.80}$$

The matrix P satisfies the following Ricatti equation:

$$P = \Phi_\pi \Phi_\pi' + \beta A_s' P A_s - (\beta A_s' P B_s + \phi_\pi \Phi_\pi) \left\{(1 + \phi_\pi^2) + \beta B_s' P B_s\right\}^{-1} (\beta B_s' P A_s + \phi_\pi \Phi_\pi'). \tag{H.81}$$