Myopia and Anchoring*

George-Marios Angeletos† **Zhen Huo**‡

September 24, 2020

Abstract

We develop an equivalence between the equilibrium effects of incomplete information and those of two behavioral distortions: myopia, or extra discounting of the future; and anchoring of current behavior to past behavior, as in models with habit persistence or adjustment costs. We show how these distortions depend on higher-order beliefs and GE mechanisms, and how they can be disciplined by evidence on expectations. We finally illustrate the use of our toolbox with a quantitative application in the context of inflation, a bridge to the HANK literature, and an extension to networks.

*We are grateful to three anonymous referees and the editor, Mikhail Golosov, for extensive feedback, and to Chris Sims, Alexander Kohlhas, Luigi Iovino, and Alok Johri for discussing our paper in, respectively, the 2018 NBER Monetary Economics meeting, the 2018 Cambridge/INET conference, the 2018 Hydra workshop, and the 2018 Canadian Macro Study Group. We also acknowledge useful comments from Jaroslav Borovicka, Simon Gilchrist, Jennifer La'O, John Leahy, Kristoffer Nimark, Stephen Morris, Mikkel Plagborg-Moller, Dmitriy Sergeyev, and seminar participants at the aforementioned conferences, BC, BU, Columbia, NYU, LSE, UCL, UBC, Chicago Fed, Minneapolis Fed, Stanford, Carleton, Edinburgh, CUHK, the 2018 Duke Macro Jamboree, the 2018 Barcelona GSE Summer Forum EGF Group, the 2018 SED Meeting, the 2018 China International Conference in Economics, and the 2018 NBER Summer Institute. Angeletos acknowledges the support of the National Science Foundation under Grant Number SES-1757198.

†MIT and NBER; angelet@mit.edu.

‡Yale University; zhen.huo@yale.edu.

1 Introduction

What are the macroeconomic effects of informational frictions? How do they depend on general equilibrium (GE) mechanisms, market structures, and agent heterogeneity? And how can they be quantified?

We develop a toolbox for addressing such questions and illustrate its use. On the theoretical front, we offer an illuminating representation result and draw connections to the literatures on networks and HANK models. On the quantitative front, we show how to extract the informational friction from survey evidence on expectations and proceed to argue that it can rationalize sizable sluggishness in the response of inflation and aggregate spending to shocks.

Framework. Our starting point is a representative-agent model, in which an endogenous outcome of interest, denoted by a_t, obeys the following law of motion:

$$a_t = \varphi\xi_t + \delta\mathbb{E}_t[a_{t+1}], \tag{1}$$

where ξ_t is the underlying stochastic impulse, or fundamental, $\varphi > 0$ and $\delta \in (0,1]$ are fixed scalars, and $\mathbb{E}_t[\cdot]$ is the rational expectation of the representative agent.

Condition (1) stylizes a variety of applications. In the textbook New Keynesian model, this condition could be either the New Keynesian Philips Curve (NKPC), with a_t standing for inflation and ξ_t for the real marginal cost, or the Euler condition of the representative consumer (a.k.a. the Dynamic IS curve), with a_t standing for aggregate spending and ξ_t for the real interest rate. Alternatively, this condition can be read as an asset-pricing equation, with ξ_t standing for the asset's dividend and a_t for its price.

We depart from these benchmarks by letting people have a noisy "understanding" of the economy, in the sense of incomplete information. The friction could be the product of dispersed knowledge (Lucas, 1972) or rational inattention (Sims, 2003). And it is the source of both first- and higher-order uncertainty. Relative to the frictionless, full-information, rational-expectations benchmark, there is therefore not only gradual learning of the exogenous innovations, but also a friction in how people reason about others (Morris and Shin, 1998; Tirole, 2015) and thereby about GE effects (Angeletos and Lian, 2018).

An Observational Equivalence. Our main result is a representation of the equilibrium effects of the informational friction in terms of two behavioral distortions. Under appropriate assumptions, the equilibrium dynamics of the aggregate outcome a_t in the incomplete-information economy are shown to coincide with that of a representative-agent economy in which condition (1) is modified as follows:

$$a_t = \varphi\xi_t + \delta\omega_f\mathbb{E}_t[a_{t+1}] + \omega_b a_{t-1}, \tag{2}$$

for some $\omega_f < 1$ and $\omega_b > 0$. The first distortion ($\omega_f < 1$) represents myopia towards the future, the second ($\omega_b > 0$) anchors current outcomes to past outcomes. One dulls the forward-looking behavior, the other adds a backward-looking element akin to habit or adjustment costs.

Crucially, both distortions increase not only with the level of noise but also with parameters that regulate the strategic interaction, or the GE feedback in the economy. Economies in which the Keynesian cross is steeper, firms are more strategic, or input-output linkages are stronger behave *as if* they are populated by more impatient and more backward-looking agents.

Underlying insights and marginal contribution. The documented effects encapsulate the role of higher-order beliefs. To fix ideas, consider the response of aggregate demand (= a_t) to a drop in the real interest rate (= ξ_t). A consumer that becomes aware of this event now may nevertheless doubt that others will be aware of the same event in the near future and may therefore also doubt that aggregate spending will go up. As this logic applies for the average consumer, the economy as a whole systematically underestimates the future movements in aggregate income, and behaves like a representative agent that excessively discounts the future. And the larger the dependence of spending on income, or the steeper the Keynesian cross, the larger this discounting.

This explains the documented myopia. The anchoring, on the other hand, has to do with learning. As more times passes since the occurrence of any given shock, consumers become progressively more aware of it. But higher-order beliefs adjust more sluggishly than first-order beliefs—equivalently, the expectations of income adjust more sluggishly than expectations of interest rates. This reduces the speed of adjustment in aggregate spending, or equivalently it increases the apparent dependence of current spending on past spending. And the steeper the Keynesian cross, the larger this effect, too.

Versions of these insights have been documented in the literature before, albeit not in the sharp form offered here.[1] Relative to the state of the art, our theoretical contribution contains: the bypassing of the curse of dimensionality in higher-order beliefs; the existence, uniqueness and analytical characterization of the equilibrium; the aforementioned observational-equivalence result; and an extension to a class of incomplete-information networks. This in turn paves the way to our applied contribution, which we detail below.

DSGE, micro to macro, and bounded rationality. Our observational equivalence offers the sharpest to-date illustration of how informational frictions may substitute for the ad hoc forms of sluggish adjustment employed in the DSGE literature: the backward-looking element in condition (2) is akin to that introduced by habit persistence in consumption, adjustment costs to investment, or indexation of prices to past inflation.

Crucially, the documented distortions increase not only with the level of noise but also with parameters that regulate the strength of GE feedback loops and the associated importance of higher-order be-

[1] In particular, the role of learning as source of sluggish adjustment in behavior is the common theme of Sims (2003) and Mankiw and Reis (2002); the higher sluggishness of higher-order beliefs relative to first-order beliefs has been emphasized by Woodford (2003) and Morris and Shin (2006); and the role of higher-order beliefs as a source of as-if myopia has been highlighted by Angeletos and Lian (2018).

liefs. In the context of the NKPC, examples of such parameters include the frequency of price adjustment, the degree of market concentration, and the input-output matrix; and in the context of the Dynamic IS curve, they include liquidity constraints and consumer heterogeneity.

Our analysis also yields the following, seemingly paradoxical, conclusion: more responsiveness at the micro level often comes together with more sluggishness at the macro level. For instance, a smaller Calvo friction maps to more sluggishness in aggregate inflation, and a higher marginal propensity to consume (MPC) maps to more habit-like persistence in aggregate consumption. In both cases, the reason is that the larger micro-level responsiveness is tied to a larger bite of higher-order uncertainty.

At the same time, our result builds a bridge to a recent literature that emphasizes how lack of common knowledge (Angeletos and Lian, 2018) and related kinds of bounded rationality (Farhi and Werning, 2019; Gabaix, 2020; Garcıa-Schmidt and Woodford, 2019) make agents behave *as if* they are myopic. But whereas this prior literature has restricted the belief error triggered by any shock to be time-invariant, our analysis lets it decay with the age of the shock, thanks to the accommodation of learning. This explains why our approach yields not only $\omega_f < 1$ but also $\omega_b > 0$, which is exactly what the data want.

Connection to evidence on expectations. Our results facilitate a simple quantitative strategy. We show how estimates of ω_f and ω_b can be obtained by combining knowledge about GE parameters with an appropriate moment of the average forecasts. Such a moment is estimated in Coibion and Gorodnichenko (2015), or CG for short: it is the the coefficient of the regression of the average forecast errors on past forecast revisions.

The basic intuition is that a higher value for this moment indicates a larger informational friction. But both the structural interpretation of this moment and its mapping to the macroeconomic dynamics is modulated by the GE feedback. When this feedback is strong enough, a modest friction by the CG metric may camouflage a large friction in terms of the values for ω_f and ω_b.

At the same time, we explain why the evidence on the under-reaction of average forecasts provided in CG is more "reliable" for our purposes than the conflicting evidence on the over-reaction of individual forecasts provided in Bordalo et al. (2020) and Broer and Kohlhas (2019). In an extension that adds a behavioral element as in those papers (a form of overconfidence), we can vary the theory's implications about individual forecasts without varying the structural relation between average forecasts and aggregate outcomes. The values of ω_f and ω_b are thus pinned down solely by the CG moment.

Applications: NKPC, HANK, and Asset Pricing. Our first application (Section 6) concerns inflation. Using our toolbox, we show that the friction implicit in surveys of expectations is large enough to rationalize existing estimates of the Hybrid NKPC. This complements Nimark (2008), which articulated the basic idea but did not discipline the theory with expectations data. To the best of our knowledge, ours is indeed the first estimate of what the available evidence of expectations means for inflation dynamics.

Echoing a core theme of our paper, we show that most of the documented effect regards the expectations of the behavior of others (inflation) rather than the expectations of the fundamental (real marginal cost). We finally put forward three ideas, all of which stem from the endogeneity of the Hybrid NKPC under the prism of our analysis. The first two draw a possible causal link from the increase in market concentration and the conduct of monetary policy to the reduction in inflation persistence. The third highlights that the economy's production network may influence not only the slope of the Philips curve (as in Rubbo, 2020; La'O and Tahbaz-Salehi, 2020) but also its backward-looking element.

Our second application (Section 7) turns to aggregate demand. As already mentioned, our theory provides a micro-foundation of habit-like persistence in aggregate spending. For a plausible calibration, this persistence is quantitatively comparable to that assumed in the DSGE literature, but requires no actual habit at the micro level. This helps reconcile the gap between the levels of habit required to match the macroeconomic time series and the much smaller levels estimated in microeconomic data (Havranek, Rusnak, and Sokolova, 2017).

Relatedly, because the as-if myopia and habit increase with the MPC, our results help reconcile the high responsiveness of consumer spending to income shocks at the micro level with the sluggishness of aggregate spending to interest-rate shocks at the macro level.[2] This hints at a link between our contribution and the emerging HANK literature. We take a step in this direction by studying a heterogeneous-agent extension of our setting and showing the following property in it: a positive cross-sectional correlation between MPC and income cyclicality, like that documented empirically in Patterson (2019), amplifies the expectations-driven sluggishness in the response of aggregate spending to monetary policy.

Other applications include investment (Appendix F) and asset pricing (Appendix G). In the latter context, our results illustrate how higher-order uncertainty may be the source of both momentum and excessive discounting. They also suggest that both distortions may be greater at the level of the entire stock market than at the level of the stock of a particular firm, which in turn may help rationalize Samuelson's dictum (Jung and Shiller, 2005).[3]

Networks. Our HANK application is an example of how our toolbox can be extended to a class of networks. In this context, we offer a tractable characterization of the equilibrium dynamics as functions of the network and information structures. This builds a bridge to a growing literature that emphasizes the network structure of the economy but often ignores informational frictions.[4]

[2] A similar point has been made recently by Auclert, Rognlie, and Straub (2020).

[3] Choi, Rondina, and Walker (2020) also attempt to rationalize the discrepancies between aggregate and individual asset prices based on incomplete information and segmented markets, but their work focuses on pricing efficiency and volatility instead of momentum and discounting.

[4] A few notable exemptions are Bergemann, Heumann, and Morris (2017) and Golub and Morris (2019) on the abstract front, and Nimark, Chahrour, and Pitschner (2019), Auclert, Rognlie, and Straub (2020) and La'O and Tahbaz-Salehi (2020) on the applied front. None of these papers, however, share either our analytical results or our emphasis on forward-looking behavior.

2 Framework

In this section we set up our framework and illustrate its applicability.

2.1 Basic Ingredients

Time is discrete, indexed by $t \in \{0,1,...\}$, and there is a continuum of agents, indexed by $i \in [0,1]$. At any t, each agent chooses an action $a_{i,t} \in \mathbb{R}$. Let a_t be the average action. Best responses admit the following recursive formulation:

$$a_{i,t} = \mathbb{E}_{i,t}\left[\varphi\xi_t + \beta a_{i,t+1} + \gamma a_{t+1}\right], \tag{3}$$

where ξ_t is an underlying fundamental, $\mathbb{E}_{i,t}[\cdot]$ is the agent's expectation in period t, and (φ,β,γ) are parameters, with $\varphi > 0$, $\gamma \in [0,\delta)$, and $\beta \equiv \delta - \gamma$, for some $\delta \in (0,1)$. As it will become clear, δ parameterizes the agent's overall concern about the future and γ the GE, or strategic, considerations.

Iterating on condition (3) yields the following representation of i's best response:

$$a_{i,t} = \sum_{k=0}^{\infty} \beta^k \mathbb{E}_{i,t}\left[\varphi\xi_{t+k}\right] + \gamma \sum_{k=0}^{\infty} \beta^k \mathbb{E}_{i,t}\left[a_{t+k+1}\right]. \tag{4}$$

While the recursive form (3) is more convenient for certain derivations, the extensive form given above is more precise because it embeds appropriate "boundary" conditions for $t \to \infty$.[5] It also makes salient how a agent's optimal behavior at any given point of time depends on her expectations of the *entire* future paths of the fundamental and of the average action.

Aggregating condition (4) yields the following equilibrium restriction:

$$a_t = \varphi \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t\left[\xi_{t+k}\right] + \gamma \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t\left[a_{t+k+1}\right], \tag{5}$$

where $\overline{\mathbb{E}}_t[.]$ denotes the average expectation in the population. This condition highlights the fixed-point relation between the equilibrium value of a_t and the expectations of it. As it will become clear, this condition also allows us to nest a variety of applications.[6]

2.2 Complete Information and Beyond

Suppose that information is complete, meaning that all agents share the *same* information and this fact itself is common knowledge. The economy then admits a representative agent. That is, $a_{i,t} = a_t$ and

[5] Namely, we have imposed that, for any date τ and history, $\lim_{t\to\infty} \beta^t \mathbb{E}_{i,\tau}[a_{i,t}] = 0$, $\lim_{t\to\infty} \beta^t \mathbb{E}_{i,\tau}[\xi_t] = 0$, and $\lim_{t\to\infty} \beta^t \mathbb{E}_{i,\tau}[a_t] = 0$. The first property can be understood as the transversality condition. The second represents a restriction on the fundamental process, trivially satisfied when ξ_t is bounded. The third represents an equilibrium refinement.

[6] The same best-response structure is assumed in Angeletos and Lian (2018). But whereas that paper considers a non-stationary setting where ξ_t is fixed at zero for all $t \neq T$, for some given $T \geq 1$, we consider a stationary setting in which ξ_t varies in all t and, in addition, there is gradual learning over time. Our framework also reminds the *static* beauty contests studied in Morris and Shin (2002), Woodford (2003), Angeletos and Pavan (2007), and Huo and Pedroni (2020). There, agents try to predict the concurrent behavior of others. Here, they try to predict the future behavior of others.

$\mathbb{E}_{i,t} = \mathbb{E}_t$, where $\mathbb{E}_t$ stands for the representative agent's expectation, and condition (3) reduces to

$$a_t = \mathbb{E}_t[\varphi\xi_t + \delta a_{t+1}]. \tag{6}$$

This may correspond to the textbook versions of the Dynamic IS and New Keynesian Philips curves, or an elementary asset-pricing equation. By the same token, the equilibrium outcome is given by

$$a_t = \varphi \sum_{h=0}^{\infty} \delta^h \mathbb{E}_t[\xi_{t+h}]. \tag{7}$$

This can be read as "inflation equals the present discounted value of real marginal costs" or "the asset's price equals the present discounted value of its dividends."

Clearly, only the composite parameter $\delta = \beta + \gamma$ enters the determination of the equilibrium outcome: its decomposition between β and γ is irrelevant. As made clear in Section 3.1 below, this underscores that the decomposition between PE and GE considerations is immaterial in this benchmark. Furthermore, the outcome is pinned down by the expectations of the fundamental alone.

These properties hold because this benchmark imposes that agents can reason about the behavior of *others* with the same ease and precision as they can reason about their *own* behavior. Conversely, introducing incomplete (differential) information and higher-order uncertainty, as we shall do momentarily, amounts to accommodating a friction in how agents reason about the behavior of others, or about GE.

2.3 Two Examples: Dynamic IS and NKPC

Before digging any further into the theory, we illustrate how our setting can nest the two building blocks of the New Keynesian model, the Dynamic IS curve and the New Keynesian Philips curve (NKPC). The familiar, log-linearized, representative-agent versions of these equations are given by, respectively,

$$c_t = \mathbb{E}_t[-\varsigma r_t + c_{t+1}] \qquad \text{and} \qquad \pi_t = \mathbb{E}_t[\kappa \mathtt{mc}_t + \chi \pi_{t+1}],$$

where c_t is aggregate consumption, r_t is the real interest rate, π_t is inflation, $\mathtt{mc}_t$ is the real marginal cost, $\varsigma > 0$ is the elasticity of intertemporal substitution, $\kappa \equiv \frac{(1-\chi\theta)(1-\theta)}{\theta}$ is the slope of the Philips curve, $\theta \in (0,1)$ is the Calvo parameter, $\chi \in (0,1)$ is the representative agent's discount factor, and $\mathbb{E}_t$ is her expectation. Clearly, both of these conditions are nested in condition (6).

Relaxing the common-knowledge foundations of the New Keynesian model along the lines of Angeletos and Lian (2018) yields the following incomplete-information extensions of these equations:

$$c_t = -\varsigma \sum_{k=0}^{\infty} \chi^k \overline{\mathbb{E}}_t[r_{t+k}] + (1-\chi) \sum_{k=1}^{\infty} \chi^{k-1} \overline{\mathbb{E}}_t[c_{t+k}], \tag{8}$$

$$\pi_t = \kappa \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t[\mathtt{mc}_{t+k}] + \chi(1-\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t[\pi_{t+k+1}], \tag{9}$$

where $\overline{\mathbb{E}}_t$ denotes the average expectation of the consumers in condition (8) and that of the firms in condition (9). The first equation is nested in condition (5) by letting $a_t = c_t$, $\xi_t = r_t$, $\varphi = -\varsigma$, $\beta = \chi$, $\gamma = 1-\chi$, and $\delta = 1$; the second by letting $a_t = \pi_t$, $\xi_t = \texttt{mc}_t$, $\varphi = \kappa$, $\beta = \chi\theta$, $\gamma = \chi(1-\theta)$ and $\delta = \chi$.

To understand condition (8), recall that the Permanent Income Hypothesis gives consumption as a function of the present discounted value of income. Incorporating variation in the real interest rate and heterogeneity in beliefs, and using the fact that aggregate income equals aggregate spending in equilibrium, yields condition (8). Finally, note that $1-\chi$ measures the marginal propensity to consume (MPC) out of income. The property that $\gamma = 1-\chi$ therefore means that, in this context, γ captures the slope of the Keynesian cross, or the GE feedback between spending and income.

To understand condition (9), recall that a firm's optimal reset price is given by the present discounted value of its nominal marginal cost. Aggregating across firms and mapping the average reset price to inflation yields condition (9). When all firms share the same, rational expectations, this condition reduces to the familiar, textbook version of the NKPC. Away from that benchmark, condition (9) reveals the precise manner in which expectations of future inflation (the behavior of firms) feed into current inflation. Note in particular that $\gamma = \chi(1-\theta)$, which means that the effective degree of strategic complementarity increases with the frequency of price adjustment. This is because the feedback from the expectations of future inflation to current inflation increases when a higher fraction of firms are able to adjust their prices today on the basis of such expectations.

3 The Equivalence Result

This section contains the core of our contribution. We motivate the requisite assumptions, solve for the rational-expectations fixed point, develop our observation-equivalence result, and discuss the main insights encapsulated in it.

3.1 Higher-Order Beliefs: The Wanted Essence and the Unwanted Complexity

Higher-order beliefs are synonymous to how agents reason about GE effects. To see this, revisit condition (5), which allows the following decomposition of the aggregate outcome:

$$a_t = \underbrace{\varphi \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[\xi_{t+k}\right]}_{\text{PE component}} + \underbrace{\gamma \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t \left[a_{t+k+1}\right]}_{\text{GE component}}. \tag{10}$$

We label the first term as the PE component because it captures the agents' response to any innovation holding constant their expectations about the endogenous outcome; the additional change triggered by any adjustment in these expectations, or the second term above, represents the GE component.

Consider now two economies, labeled A and B, that share the same $\delta \equiv \beta + \gamma$ but have a different mixture of β and γ. Economy A features $\beta = \delta$ and $\gamma = 0$, which means that GE considerations are entirely absent. Economy B features $\beta = 0$ and $\gamma = \delta$, which corresponds to "maximal" GE considerations.

In economy A, condition (5) reduces to $a_t = \varphi \sum_{k=0}^{\infty} \delta^k \overline{\mathbb{E}}_t[\xi_{t+k}]$, that is, only the first-order beliefs of the fundamental matter. This is similar to the representative-agent benchmark, except that the representative agent's expectations are replaced by the average expectations in the population. In economy B, instead, condition (5) reduces to $a_t = \varphi \overline{\mathbb{E}}_t[\xi_t] + \delta \overline{\mathbb{E}}_t[a_{t+1}]$ and recursive iteration yields

$$a_t = \varphi \sum_{h=1}^{\infty} \delta^h \overline{\mathbb{F}}_t^h [\xi_{t+h-1}], \tag{11}$$

where, for any variable X, $\overline{\mathbb{F}}_t^1[X] \equiv \overline{\mathbb{E}}_t[X]$ denotes the average first-order forecast of X and, for all $h \geq 2$, $\overline{\mathbb{F}}_t^h[X] \equiv \overline{\mathbb{E}}_t\left[\overline{\mathbb{F}}_{t+1}^{h-1}[X]\right]$ denotes the corresponding h-th order forecast. The key difference from both the representative-agent benchmark and economy A is the emergence of such higher-order beliefs. These represent GE considerations, or the agents' reasoning about the behavior of others.

The logic extends to the general case, in which both β and γ are positive. The only twist is that the relevant set of higher-order beliefs is significantly richer than that seen in condition (11). Indeed, let $\zeta_t \equiv \sum_{\tau=0}^{\infty} \beta^\tau \xi_{t+\tau}$ and consider the following set of forward-looking, higher-order beliefs:

$$\overline{\mathbb{E}}_{t_1}[\overline{\mathbb{E}}_{t_2}[\cdots[\overline{\mathbb{E}}_{t_h}[\zeta_{t+k}]\cdots]],$$

for any $t \geq 0$, $k \geq 2$, $h \in \{2, ..., k\}$, and $\{t_1, t_2, ..., t_h\}$ such that $t = t_1 < t_2 < ... < t_h = t + k$. As behavior depends on all these higher-order beliefs, this adds considerable complexity relative to the $\beta = 0$ case. For instance, when $k = 10$ (thinking about the outcome 10 periods later), there are 210 beliefs of the fourth order that are relevant when $\beta > 0$ compared to only one such belief when $\beta = 0$.[7]

An integral part of our contribution is the bypassing of this complexity. The assumptions that permit this bypassing are spelled out below. They come at the cost of some generality, in particular we abstract from the possible endogeneity of information.[8] But they also bear significant gains on both the theoretical and the quantitative front, which will become evident as we proceed.

3.2 Specification

We henceforth make two assumptions. First, we let the fundamental ξ_t follow an AR(1) process:

$$\xi_t = \rho \xi_{t-1} + \eta_t = \frac{1}{1 - \rho L} \eta_t, \tag{12}$$

where $\eta_t \sim \mathcal{N}(0,1)$ is the period-t innovation, L is the lag operator, and $\rho \in (0,1)$ parameterizes the persistence of the fundamental. Second, we assume that agent i receives a new private signal in each

[7] More generally, for any t and any $k \geq 2$, there are now $k-1$ types of second-order beliefs, plus $(k-1) \times (k-2)/2$ types of third-order beliefs, and so on.

[8] This abstraction is the right benchmark for our purposes, including the connections built to the evidence on expectations: this evidence helps discipline the theoretical mechanisms we are concerned with, but contains little guidance on the degree or manner in which information may be endogenous.

period t, given by

$$x_{i,t} = \xi_t + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}(0, \sigma^2), \tag{13}$$

where $\sigma \geq 0$ parameterizes the informational friction (the level of noise). The agent's information in period t is the history of signals up to that period.

As anticipated in the previous subsection, these assumptions aim at minimizing complexity without sacrificing essence. Borrowing from the literature on rational inattention, we also invite a flexible interpretation of our setting as one where fundamentals and outcomes are observable but cognitive limitations makes agents act *as if* they observe the entire state of nature with idiosyncratic noise. But instead of endogenizing the noise, we fix it in a way that best serves our purposes.

3.3 Solving the Rational-Expectations Fixed Point

Consider momentarily the frictionless benchmark ($\sigma = 0$), in which case the outcome is pinned down by first-order beliefs, as in condition (7). Thanks to the AR(1) specification for the fundamental, we have $\mathbb{E}_t[\xi_{t+k}] = \rho^k \xi_t$, for all $t, k \geq 0$. We thus reach the following result, which states that the complete-information outcome follows the same, up to a rescaling, AR(1) process as the fundamental.[9]

Proposition 1. *In the frictionless benchmark* ($\sigma = 0$)*, the equilibrium outcome is given by*

$$a_t = a_t^* \equiv \frac{\varphi}{1-\rho\delta}\xi_t = \frac{\varphi}{1-\rho\delta}\frac{1}{1-\rho L}\eta_t. \tag{14}$$

Consider next the case in which information is incomplete ($\sigma > 0$). As already explained, the outcome is then a function of an infinite number of higher-order beliefs. Despite the assumptions made here about the process of ξ_t and the signals, these beliefs remain exceedingly complex.

Let us illustrate this point. Using the Kalman filter, one can readily show that the first-order belief $\overline{\mathbb{E}}_t[\xi_t]$ obeys the following AR(2) dynamics:

$$\overline{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{\lambda}{\rho}\right)\left(\frac{1}{1-\lambda L}\right)\xi_t, \tag{15}$$

where $\lambda = \rho(1-g)$ and $g \in (0,1)$ is the Kalman gain, itself a decreasing function of the level of noise.[10] It follows that the second-order belief $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_{t+1}[\xi_{t+1}]]$ follows an ARMA(3,1). By induction, for any $h \geq 1$, the h-th order belief $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_{t+1}[...\overline{\mathbb{E}}_{t+h}[\xi_{t+h}]]$ follows an ARMA$(h+1, h-1)$. Beliefs of higher order thus exhibit increasingly complex dynamics.

As explained in Section 3.1, the above set of higher-order beliefs is the relevant one when $\beta = 0$. The general case with $\beta > 0$ is subject to an even greater curse of dimensionality in terms of higher-order beliefs. And yet, this complexity vanishes once we focus on the rational-expectations fixed point:

[9] All proofs are delegated to Appendix A.

[10] The Kalman gain is given by the unique $g \in (0,1)$ such as that $(1-g) = (1-\rho^2(1-g))g\sigma^2$. This yields g as a continuous and decreasing function of σ, with $g = 1$ when $\sigma = 0$ and $g \to 0$ when $\sigma \to \infty$.

under our assumptions, the fixed point turns out to be merely an AR(2) process, whose exact form is characterized below.

Proposition 2 (Solution)**.** *The equilibrium exists, is unique and is such that the aggregate outcome obeys the following law of motion:*

$$a_t = \left(1 - \frac{\vartheta}{\rho}\right)\left(\frac{1}{1-\vartheta L}\right) a_t^*, \tag{16}$$

where a_t^* is the frictionless counterpart, obtained in Proposition 1, and where ϑ is a scalar that satisfies $\vartheta \in (0,\rho)$ and that is given by the reciprocal of the largest root of the following cubic:

$$C(z) \equiv -z^3 + \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + (\delta-\gamma)\right) z^2 \tag{17}$$
$$- \left(1 + (\delta-\gamma)\left(\rho + \frac{1}{\rho}\right) + \frac{\delta}{\rho\sigma^2}\right) z + (\delta-\gamma).$$

Condition (16) gives the incomplete-information dynamics as a transformation of the complete-information counterpart. This transformation is indexed by ϑ. Relative to the frictionless benchmark (herein nested by $\vartheta = 0$), a higher ϑ means both a smaller impact effect, captured by the factor $1 - \frac{\vartheta}{\rho}$ in condition (16), and a more sluggish build up over time, captured by the lag term ϑL.

To understand the math behind the result, let $\beta = 0$ momentarily. In this case, the outcome obeys

$$a_t = \varphi \overline{\mathbb{E}}_t[\xi_t] + \gamma \overline{\mathbb{E}}_t[a_{t+1}]. \tag{18}$$

If we guess that a_t follows an AR(2), we have that $\overline{\mathbb{E}}_t[a_{t+1}]$ follows an ARMA(3,1). As already noted, $\overline{\mathbb{E}}_t[\xi_t]$ follows the AR(2) given in (15). The right-hand side of the above equation is therefore the sum of an AR(2) and an ARMA(3,1). If the latter was arbitrary, this sum would have returned an ARMA(5,3), contradicting our guess that a_t follows an AR(2). But the relevant ARMA(3,1) is *not* arbitrary.

Because the impulse behind a_t is ξ_t, one can safely guess that a_t inherits the root of ξ_t. That is, $(1-\vartheta L)(1-\rho L)a_t = b\eta_t$, for some scalars b and ϑ. This in turn implies that the AR roots of the ARMA(3,1) process for $\overline{\mathbb{E}}_t[a_{t+1}]$ are the reciprocals of ρ, ϑ and λ. As seen in (15), the roots of $\overline{\mathbb{E}}_t[\xi_t]$ are the reciprocals of ρ and λ. It follows that the sum in the right-hand side of (18) is at most an ARMA(3,1) of the following form:

$$a_t = \frac{c(1-dL)}{(1-\vartheta L)(1-\rho L)(1-\lambda L)}\eta_t, \tag{19}$$

where c and d are functions of b and ϑ. For our guess to be correct, it has to be that $d = \lambda$ and $c = b$. The first equation, which lets the MA part and the last AR part cancel out so as to reduce the above to an AR(2), and yields (17). The second equation, which pins down the scale, yields $b = \left(1 - \frac{\vartheta}{\rho}\right)\left(\frac{\varphi}{1-\rho\delta}\right)$.

This is the crux of how the rational expectations fixed point works. The proof presented in Appendix A follows a somewhat different path, which is more constructive, accommodates $\beta > 0$, and can be extended to richer settings along the lines of Huo and Takayama (2018).

When $\gamma = 0$, GE considerations are absent, the outcome is pinned down by first-order beliefs, and

Proposition 2 holds with $\vartheta = \lambda$, where λ is the same root as that seen in (15). When instead $\gamma > 0$, GE considerations and higher-order beliefs come into play. As already noted, these beliefs follow complicated ARMA processes of ever increasing orders. And yet, the equilibrium continues to follow an AR(2) process. The only twist is that $\vartheta > \lambda$, which, as mentioned above, means that the equilibrium outcome exhibits less amplitude and more persistence than the first-order beliefs. This is the empirical footprint of higher-order uncertainty, or of the kind of imperfect GE reasoning accommodated in our analysis.

Below, we translate these properties in terms of our observational-equivalence result (Propositions 3 and 5). The following corollary, which proves useful when connecting the theory to evidence on expectations, is also immediate.

Corollary 1 (Forecasts)**.** *Any moment of the joint process of the aggregate outcome, a_t, and of the average forecasts, $\overline{\mathbb{E}}_t[a_{t+k}]$ for all $k \geq 1$, are functions of only the triplet $(\vartheta, \lambda, \rho)$, or equivalently of $(\gamma, \delta, \rho, \sigma)$.*

3.4 The Equivalence Result

Momentarily put aside our incomplete-information economy and, instead, consider a "behavioral" economy populated by a representative agent whose aggregate Euler condition (6) is as follows:

$$a_t = \varphi \xi_t + \delta \omega_f \mathbb{E}_t [a_{t+1}] + \omega_b a_{t-1}, \tag{20}$$

for some scalars ω_f, ω_b. It is easy to verify that the equilibrium process of a_t in this economy is an AR(2) whose coefficients are functions of (ω_f, ω_b) and (φ, δ, ρ). In comparison, the equilibrium process of a_t in our incomplete-information economy is an AR(2) whose coefficients determined as in Proposition 2. Matching the coefficients of the two AR(2) processes, and characterizing the mapping from the latter to the former, we reach the following result.

Proposition 3 (Observational Equivalence)**.** *Fix $(\varphi, \delta, \gamma, \rho)$. For any noise level $\sigma > 0$ in the incomplete-information economy, there exists a unique pair (ω_f, ω_b) in the behavioral economy such that the two economies generate the same joint dynamics for the fundamental and the aggregate outcome. Furthermore, this pair satisfies $\omega_f < 1$ and $\omega_b > 0$.*

This result allows one to recast the informational friction as the combination of two behavioral distortions: extra discounting of the future, or myopia, in the form of $\omega_f < 1$; and backward-looking behavior, or anchoring of the current outcome to past outcome, in the form of $\omega_b > 0$.

This representation is, of course, equivalent to the closed-form solution provided in Proposition 2. We prefer the new representation not only because it serves the applied purposes of our paper, but also because the main insights about myopia and anchoring extend to richer settings, while the specific AR(2) solution provided in Proposition 2 does not. This idea is formalized in Appendix H.

3.5 The Roles of Noise and GE Considerations

As one would expect, both distortions increase with the level of noise.

Proposition 4 (Noise)**.** *A higher σ maps to a lower ω_f and a higher ω_b.*

What this result, however, fails to highlight is the dual meaning of "noise" in our setting: a higher σ represents not only less accurate information about the fundamental (larger first-order uncertainty) but also more friction in how agents reason about others (larger higher-order uncertainty). The latter, strategic or GE, channel is highlighted by the next result.

Proposition 5 (GE)**.** *Consider an increase in the relative importance of GE considerations, as captured by an increase in γ holding $\delta \equiv \beta + \gamma$, as well as σ and ρ, constant. This maps to both greater myopia (lower ω_f) and greater anchoring (higher ω_b).*

This result circles backs to our discussion in Section 3.1 regarding the interpretation of higher-order uncertainty as a distortion in agents' GE reasoning. It also anticipates a point we make in Section 5. While the kind of evidence on informational frictions provided by Coibion and Gorodnichenko (2015) is an essential ingredient for the quantitative evaluation of the assumed friction, it is not sufficient. One must combine such evidence with knowledge of how important the GE feedback from expectations to actual behavior is.

3.6 Robustness

The results presented above depend on stark assumptions about the process of ξ_t and the information structure. But the key insights regarding myopia, anchoring, and the role of higher-order beliefs are more general. Appendix H shows how to generalize these insights in a setting that allows ξ_t to follow an essentially arbitrary MA process, as well as information to diffuse in a flexible manner.[11] The elegance of our observational-equivalence result is lost, but the essence remains.

Another extension, better suited for applied purposes, is offered in Section 8. There, we consider a multi-variate analogue of condition (4). This allows one to handle the full, three-equation New Keynesian model, the HANK variant considered in Section 7, and a large class of linear networks.

4 Connection to DSGE, Bounded Rationality, and Beyond

In the end of Section 2 we sketched how our framework nests incomplete-information extensions of the Dynamic IS curve and the NKPC. We also discussed how γ relates to the slope of the Keynesian cross,

[11] Such richness is prohibitive in general. We cut the Gordian knot by orthogonalizing the information about the innovations occurring at different points of time. Although this modeling approach is unusual, it nests "sticky information" (Mankiw and Reis, 2002) as a special case and clarifies the theoretical mechanisms.

or the income-spending multiplier, in the first context and to the frequency of price adjustment in the second. The following translations of our abstract results are thus immediate.

Corollary 2. *Applying our result to condition (9) yields the following NKPC:*

$$\pi_t = \kappa mc_t + \omega_f \chi \mathbb{E}_t[\pi_{t+1}] + \omega_b \pi_{t-1}. \tag{21}$$

In this context, the distortions increase with the frequency of price adjustment.

Corollary 3. *Applying our result to condition (8) yields the following Dynamic IS curve:*

$$c_t = -\varsigma r_t + \omega_f \mathbb{E}_t[c_{t+1}] + \omega_b c_{t-1}. \tag{22}$$

In this context, the distortions increase with the MPC, or the slope of the Keynesian cross.

Condition (21) looks like the Hybrid NKPC. Condition (22) looks like the Euler condition of representative consumer who exhibits habit persistence plus myopia. Appendix F offers a related result for investment: we take a model in which adjustments cost depend on the change in the stock of capital, as in traditional Q theory; add incomplete information; and show that this model looks like a model in which adjustment costs depend on the change in the rate of investment.

Together, these results illustrate how informational frictions can substitute for the more ad hoc sources of sluggishness in *all* the equations of DSGE models. The basic idea is familiar from previous works (e.g., Sims, 2003; Mankiw and Reis, 2002; Woodford, 2003; Nimark, 2008). The added value here is the sharpness of the provided representation and the following, complementary lessons.

First, we build a bridge to a recent literature that shows how lack of common knowledge and related forms of bounded rationality make agents behave *as if* they are myopic. These works help generate $\omega_f < 1$ but restrict $\omega_b = 0$. In Angeletos and Lian (2018), this is because there is no learning. In Farhi and Werning (2019), Garcıa-Schmidt and Woodford (2019) and Iovino and Sergeyev (2017), it is a direct implication of the adopted solution concept: level-k thinking amounts to equating beliefs of order $h \leq k$ to their complete-information counterparts, and beliefs of order $h > k$ to zero. This makes agents underestimate GE effects, which maps to $\omega_f < 1$, but precludes the mistake in beliefs to be corrected over time, which maps to $\omega_b = 0$. Our approach, instead, naturally delivers both $\omega_f < 1$ and $\omega_b > 0$, which is what the macroeconomic data want.[12] By the same token, our approach allow both for under-reaction and momentum in average expectations, which is what the available survey evidence want.

Second, we offer a new rationale for why the information-driven sluggishness may loom large at the macro level even if is absent at the micro level. Previous work has emphasized that agents may naturally have less information about aggregate shocks than about idiosyncratic shocks (Maćkowiak and

[12]This point applies to dynamic settings. In static games such as Morris and Shin (2002), the three approaches are observationally equivalent vis-a-vis the macroeconomic time series.

Wiederholt, 2009). We add that higher-order uncertainty effectively amplifies the friction at the macro level. We further clarify these points in Appendix C by considering an extension of our framework with idiosyncratic shocks. And in Appendix G, we discuss how the exact same logic transported to an asset-pricing context may help rationalize larger momentum at the macro level than at the micro level, or what is known as Samuelson's dictum (Jung and Shiller, 2005).

Third, by tying the macro-level distortions to strategic complementarity and GE feedbacks, we highlight how the former can be endogenous to market structures and policies that regulate the latter. We come back to this point in Section 6.

Fourth, in the context of the NKPC, we show that higher price flexibility contributes to more sluggishness in inflation by intensifying the role of higher-order beliefs. This seems an intriguing, new addition to the "paradoxes of flexibility." And in the context of the Dynamic IS curve, we tie the habit-like persistence in consumption to the MPC, or the slope of the Keynesian cross. This hints at the promise of incorporating incomplete information in the HANK literature, an idea we expand on in Section 7.

Finally, we offer a simple strategy for quantifying the distortions of interest. We spell out the elements of this strategy in the next section and put it at work in our subsequent applications to inflation and consumption dynamics.

5 Connection to Evidence on Expectations

Proposition 3 ties the documented distortions to σ. This parameter may not be a priori known to the analyst ("econometrician"). Surveys of expectations, however, can help identify it. In this section, we use our results to map readily available evidence on expectations to the macroeconomic distortions of interest. We also clarify which subset of such evidence is best suited for our purposes (moments of average forecasts) and provide two examples of robustness for the offered mapping (one regarding overconfidence and another regarding public signals).

5.1 Calibrating the Friction

Consider Coibion and Gorodnichenko (2015), or CG for short. This paper runs the following regression on data from the Survey of Professional Forecasters:

$$a_{t+k} - \overline{\mathbb{E}}_t[a_{t+k}] = K_{\mathrm{CG}}\left(\overline{\mathbb{E}}_t[a_{t+k}] - \overline{\mathbb{E}}_{t-1}[a_{t+k}]\right) + v_{t+k,t}, \tag{23}$$

where a_t is an economic outcome such as inflation and $\overline{\mathbb{E}}_t[a_{t+k}]$ is the average ("consensus") forecast of the value of this outcome k periods later. CG's main finding is that K_{CG}, the coefficient of the above regression, is positive. That is, a positive revision in the average forecast between $t-1$ and t predicts a positive average forecast error at t.

What does this mean under the lenses of the theory? Insofar as agents are rational, an agent's forecast error ought to be orthogonal to his *own* past revision, itself an element of the agent's information set. But this does not have to be true at the aggregate level, because the past average revision may not be commonly known. More succinctly, $K_{\text{CG}} \neq 0$ is possible because the forecast error of one agent can be predictable by the past information of another agent.

Furthermore, because forecasts adjust sluggishly towards the truth, the theory suggests that K_{CG} ought to be positive and increasing in the informational friction. To illustrate this, CG treat a_t as an exogenous AR(1) process, assume the same Gaussian signals as we do, and show that in this case $K_{\text{CG}} = \frac{1-g}{g}$, where $g \in (0,1)$ is the Kalman gain, itself a decreasing function of σ. They therefore argue that their estimate of K_{CG} offers a measure of the informational friction.

In our context, a_t is endogenous to expectations. This complicates the structural interpretation and use of this measure. The level of noise now influences not only the agents' forecasting of a_t, but also its own stochastic process. Furthermore, because the level of noise interacts with the GE feedback in shaping the process for a_t, the GE parameter γ enters the mapping between σ and K_{CG}. The next result shows what exactly is going on.

Proposition 6 (K_{CG})**.** *The theoretical counterpart of the coefficient of regression (23) for $k = 1$ is given by*

$$K_{CG} = \lambda \frac{\vartheta + \rho - \rho\vartheta(\lambda + \vartheta) - \rho\lambda\vartheta(1 - \lambda\vartheta)}{(\rho - \lambda)(1 - \lambda\vartheta)(\rho + \vartheta - \lambda\rho\vartheta)}, \tag{24}$$

where λ and ϑ are defined as in Section 3.3. It follows that

(i) K_{CG} is increasing in σ, the level of noise; and

(ii) K_{CG} is decreasing in γ, the GE feedback.

The formula for K_{CG} is not particularly intuitive. However, in combination with our closed-form characterizations for λ and ϑ, it allows us to prove the two illuminating comparative statics stated above. The first verifies that CG's logic that a high value for K_{CG} signals a high informational friction extends from their PE context, where a_t follows an exogenous process, to our GE context, where the process for a_t is influenced by the informational friction. The second comparative static highlights the limits of this logic: a small value for K_{CG} could conceal a large value for σ if the GE feedback is large enough.

At first glance, this may appear to contradict our result in Proposition 5 that a higher γ translates to larger distortions in the equilibrium dynamics. But the underlying logic for both results is actually the same. When γ is higher, agents are more willing to coordinate their behavior. This reduces the reliance of behavior on private information and increases the reliance on the prior or higher-order beliefs. As this happens, the equilibrium outcome becomes less responsive to innovations. But precisely because of this reason, the reliance of the forecasts of the outcome on private information is also reduced, which means that the forecast error of one agent is less predictable by the information of another agent, and hence that the K_{CG} coefficient is closer to zero.

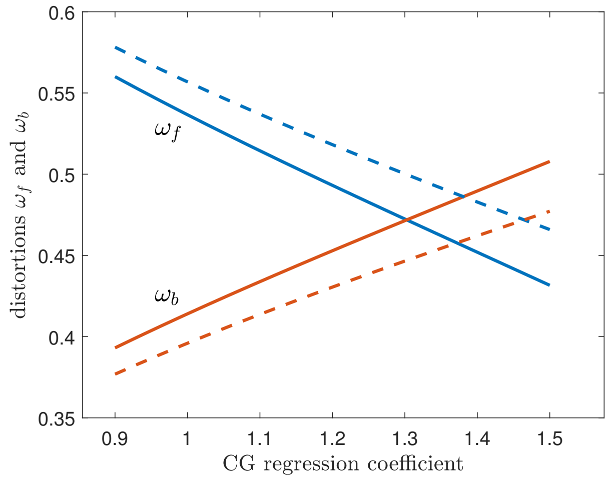

Figure 1: Myopia and Anchoring

Note: The distortions as functions of the proxy offered in Coibion and Gorodnichenko (2015). The solid lines correspond to a stronger degree of strategic complementarity, or GE feedback, than the dashed one.

What does this mean for the structural interpretation and use of the available expectations evidence? When the GE effect increases, both of the aforementioned channels work in the same direction: for given σ, a higher γ means both larger distortions in terms of (ω_f, ω_b) and a smaller observable footprint in terms of K_{CG}. The following is therefore true:

Corollary 4. *As γ increases, the same value for K_{CG} maps to both more myopia (smaller ω_f) and more anchoring (larger ω_b) in the aggregate outcome.*

This is illustrated in Figure 1. On the horizontal axis, we vary the value of K_{CG} that may be recovered from running regression (23) on the applicable expectations data. On the vertical axis, we report the predicted values for ω_f and ω_b. For given γ, a higher K_{CG} maps to a higher σ and thereby to larger distortions. But a higher γ maps to larger distortions for given K_{CG} not only because it amplifies the effect of noise, but also because it requires a larger σ to match the given K_{CG}.

5.2 Individual Forecasts and Overconfidence

So far, we have emphasized how one could make use of the moment estimated in CG, along with our tools, to obtain an estimate of ω_f and ω_b. Other moments of the *average* forecasts, such as the persistence of the average forecast errors estimated in Coibion and Gorodnichenko (2012), could serve a similar role. But what about moments of the *individual* forecasts? We next explain why such moments can be largely ignored for our purposes (but not for other purposes).

Consider, in particular, the individual-level counterpart of the CG regression, that is, the regression of one's forecast errors on one's *own* past revisions. As noted earlier, rational expectations requires that the coefficient of this regression be zero. Bordalo et al. (2020) and Broer and Kohlhas (2019) argue that this coefficient is negative in the data, supporting the presence of overconfidence. Our own take is that the evidence is inconclusive: the relevant coefficient switches signs across variables and samples (inflation vs. unemployment, pre- vs post-Volker, etc), making it hard to reject rational expectations. But even if we take for granted those papers' preposition of systematic bias in beliefs, this does not necessarily upset either our theoretical results or our proposed quantitative strategy.

We illustrate this point by augmenting our model with the same kind of over-confidence as Broer and Kohlhas (2019): whereas the actual level of noise is σ, agents perceive it to be $\hat{\sigma}$, for some $\hat{\sigma} < \sigma$. (For completeness, under-confidence, or $\hat{\sigma} > \sigma$, is also allowed.) In this extension, the gap between $\hat{\sigma}$ and σ, or the degree of overconfidence, emerges as the essential determinant of the aforementioned individual-level moment.[13] But this moment and its determinant "drop out of the picture" for our purposes:

Proposition 7. *Propositions 2–6 and Corollary 1 remain valid, modulo the replacement of σ with $\hat{\sigma}$ throughout. By implication, the mapping from K_{CG} to (ω_f, ω_b) is invariant to the degree of overconfidence.*

To understand this result, note that the perceived $\hat{\sigma}$ alone determines how much each agent's beliefs and choices vary with his information, and thereby how much the corresponding aggregates vary with the underlying fundamental. The true σ instead determines how *unequal* beliefs and choices are in the cross section, but such inequality does not matter for aggregates in our class of economies. It follows that all our results, including the characterization of (ω_f, ω_b) and K_{CG}, carry over by replacing σ with $\hat{\sigma}$.

Suppose, now, that the analyst knows all parameters except $\hat{\sigma}$ and σ and wishes to quantify the equilibrium effects of the friction under consideration (as we do, for example, in Section 6). Suppose further that the analyst combines the CG coefficient with the individual-level counterpart estimated in Bordalo et al. (2020) and Broer and Kohlhas (2019). Then, the CG coefficient *alone* allows the identification of $\hat{\sigma}$ and the quantification of its effect on the actual dynamics. The individual-level counterpart allows the identification of σ, but this does not affect the aforementioned quantitative evaluation.

A similar point applies to the cross-sectional dispersion of forecasts. A large part of it is accounted by individual-specific fixed effects, which themselves correlate with life-time experiences unrelated to the current macroeconomic context (Malmendier and Nagel, 2016). This can be accommodated in the theory by letting each agent i have a different prior mean, μ_i, about ξ_t. Such prior-mean heterogeneity is then a key determinant of the cross-sectional dispersion of forecasts. But it does not matter at all for our observational equivalence result and the offered mapping from K_{CG} to (ω_f, ω_b).

This also anticipates the exercise conducted in Table 1: for our quantitative application to inflation,

[13] Broer and Kohlhas (2019) establish this point in a setting where a_t follows an exogenous AR(1) process, but the logic extends to our context. When agents are overconfident ($\hat{\sigma} < \sigma$), they over-react to their information relative to what a rational agent would do, so a positive forecast revision today predicts a negative forecast error in the future. And the converse is true if agents are under-confident ($\hat{\sigma} > \sigma$). Also note that, although the formulation used in Bordalo et al. (2020) has different methodological underpinnings, it works in essentially the same way as the form overconfidence considered here.

we test the ability of our model to capture the cross-sectional dispersion of the forecast *errors* or the forecast *revisions* precisely because these objects partial out individual fixed effects such as those associated with the aforementioned kind of heterogeneity.

More challenging is the evidence presented in Kohlhas and Walther (2019). In direct contradiction to CG's message, these authors argue that expectations *over-react* in the sense that average forecasts errors are negatively correlated with past outcomes. They then proceed to offer a resolution based on asymmetric attention to pro-cyclical and counter-cyclical components of the forecasted variable. In Appendix I, we explain how our methods can be adapted to their setting. And in Angeletos, Huo, and Sastry (2020), we propose an alternative resolution, one based on the combination of informational frictions and over-extrapolation. But we leave this issue out of the present paper.

5.3 Public Information

So far we have have let agents observe only private signals. If we add public signals, the CG moment is no more sufficient for uniquely identifying the information structure: there are multiple combinations of the precisions of the private and public signals that generate the same value for K_{CG}. By the same token, any given value for K_{CG} maps to a set of possible values for the pair (ω_f, ω_b).

At first glance, this poses a challenge for the quantitative strategy proposed in this section. However, as explained in Appendix B and illustrated in our application to inflation below, this challenge is resolved by two key observations.

First, K_{CG} puts a tight upper bound on the relative precision of the public signal. Intuitively, as information gets more and more correlated, everybody's expectations converge to those of a representative agent, and K_{CG} converges to zero. A high value for K_{CG} therefore means *necessarily* either that there is little public information to start with, or that people pay little attention to it.

Second, by varying the precision of public information between zero and the aforementioned upper bound, we can span the entire range of values (ω_f, ω_b) that are consistent with any given value of K_{CG}. In Appendix D.3, we implement this strategy in our application to inflation, which is the topic of the next section, and show that the distortions reported therein under the simplifying assumption of no public information represent a lower bound on the distortions obtained when public information is added.

6 Application to Inflation

We now apply our toolbox the context of inflation. We argue that the theory can not only rationalize existing estimates of the Hybrid NKPC with some level of noise, but also do so with a level of noise consistent with that inferred from CG's evidence on expectations. We also illustrate how our theory ties the coefficients of the Hybrid NKPC to policy and market structures.[14]

6.1 Operationalizing the Theory

Consider the incomplete-information NKPC introduced in Section 2:[15]

$$\pi_t = \kappa \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t [\mathtt{mc}_{t+k}] + \chi(1-\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t [\pi_{t+k+1}], \tag{25}$$

Unlike the representation obtained in Corollary 2, this equation is structural: it is invariant to the process for the real marginal cost and the specification of information. But it is also hard to implement empirically, because it requires data on the term structure of the relevant forecasts over long horizons. This is where our toolbox comes handy: using our results, we can connect the above structural equation both to existing estimates of the Hybrid NKPC and to the available evidence on expectations.

To evaluate these connections, we henceforth interpret the time period as a quarter and impose the following parameterization: $\chi = 0.99$, $\theta = 0.6$, and $\rho = 0.95$. The value of χ requires no discussion. The value of θ is in line with micro data and textbook treatments of the NKPC. The value of ρ is obtained by estimating an AR(1) process on the labor share, the empirical proxy for the real marginal cost used in, inter alia, Galí and Gertler (1999) and Galí, Gertler, and Lopez-Salido (2005).[16] Finally, the value of κ is left undetermined: because this parameter scales up and down the inflation dynamics equally under any information structure, it is irrelevant for the conclusions drawn below.[17]

6.2 Connecting to Existing Estimates of the Hybrid NKPC

While an unrestricted estimation of the Hybrid NKPC allows ω_f and ω_b to be free, our theory ties them together: a higher ω_b can be obtained only if the noise is larger, which in turns requires ω_f to be smaller. A quick test of the theory is therefore whether the existing estimates of the Hybrid NKPC happen to satisfy

[14] Nimark (2008) foresaw the first part of our application by showing that an econometrician would estimate a Hybrid NKPC on artificial data generated by his model. Relative to that paper, we offer a sharper illustration of this possibility and, most importantly, let the evidence on expectations bear on the theory. Such a connection to the expectations evidence is also absent from Woodford (2003), Mankiw and Reis (2002), Reis (2006), Kiley (2007), Maćkowiak and Wiederholt (2009, 2015) and Matejka (2016). Melosi (2016) utilizes expectations data but studies a different issue, the signaling role of monetary policy. Finally, the literature on adaptive learning (Sargent, 1993; Evans and Honkapohja, 2012) also allows for the anchoring of current outcomes to past outcomes; see in particular Carvalho et al. (2017) for an application to inflation. But the anchoring found in our paper has three distinct qualities: it is consistent with rational expectations; it is tied to the strength of the GE feedback; and it is directly comparable to that found in the DSGE literature.

[15] Recall that π_t is the inflation rate, $\mathtt{mc}_t$ is the real marginal cost, $\chi \in (0,1)$ is the discount factor, $\theta \in (0,1)$ is the Calvo parameter, and $\kappa > 0$ is the slope of the NKPC. Appendix D.1 contains a detailed derivation, a discussion of the underlying assumptions, and an explanation of a mistake in versions of this condition found in some prior work.

[16] We use seasonally adjusted business sector labor share as proxy for the real marginal cost, from 1947Q1 to 2019Q2. This yields an estimate of ρ equal to 0.97 or 0.92 depending on whether we exclude or include a linear trend.

[17] In the textbook version of the NKPC, κ is itself pinned down by χ and θ. But the literature has provided multiple rationales for why κ can differ from its textbook value (e.g., it can vary with the curvature of "Kimball aggregator"). For our purposes, this amounts to treating κ as a free parameter.

this restriction. We implement this test in Figure 2. The negatively slopped line depicts the aforementioned restriction. The crosses represent the three main estimates of the pair (ω_f, ω_b) from Galí, Gertler, and Lopez-Salido (2005), and the surrounding disks give the corresponding confidence regions.[18]

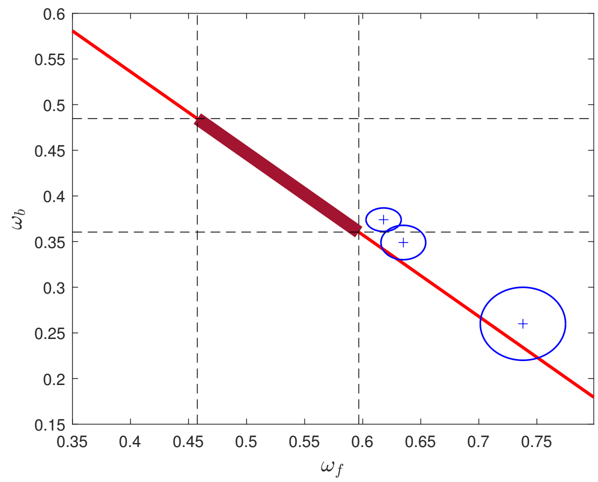

Figure 2: Testing the Theory

Note: The straight line represents the relation between ω_f and ω_b implied by the theory. Raising the level of noise maps to moving northwest along this line. The darker, thicker segment of this line corresponds to the confidence interval of K_{CG}, the relevant moment of the inflation forecasts, as reported in column (1) of Table 1 Coibion and Gorodnichenko (2015). The three crosses represent the three estimates of the pair (ω_f, ω_b) provided in Table 1 of Galí, Gertler, and Lopez-Salido (2005), and the surrounding disks give the corresponding confidence regions.

As evident in the figure, the theory passes the aforementioned test: the existing estimates of the Hybrid NKPC can be rationalized by some level of noise.[19] But is the requisite level of noise empirically plausible? We address this question next by making use of the mapping developed in Section 5.

6.3 Bringing in the Evidence on Expectations

As already noted, CG have run regression (23) using data from the Survey of Professional Forecasters.[20] Their main OLS specification, reported in column (1) of Table 1 of that paper, yields a mean estimate for

[18]The three estimates are taken from Table 1 of that paper. In particular, the left one of the three points shown in Figure 2 corresponds to $(\omega_f, \omega_b) = (0.618, 0.374)$ and is obtained by the GMM estimation of the closed-form solution that expresses current inflation as the sum of past inflation and all the expected future real marginal costs. The middle point corresponds to $(\omega_f, \omega_b) = (0.635, 0.349)$ and is obtained by GMM estimation of the hybrid NKPC directly. Finally, the right point corresponds to $(\omega_f, \omega_b) = (0.738, 0.260)$ and is obtained by a nonlinear instrumental variable estimation.

[19]Mavroeidis, Plagborg-Møller, and Stock (2014) review the extensive literature on the empirical literature of the NKPC and questions the robustness of the estimates provided by Galí, Gertler, and Lopez-Salido (2005). This debate is beyond the scope of our paper. In any event, the exercise conducted next bypasses the estimation of the Hybrid NKPC on macroeconomic data and instead infers it by calibrating our theory to survey data on expectations.

[20]In the present context, it would be preferable to have an estimate of K_{CG} for the average forecasts of a representative sample of US firms. Such an estimate is lacking in the literature, but the evidence in Coibion and Gorodnichenko (2012) suggests that the friction among firms and consumers is, as one would expect, larger than that among professional forecasters.

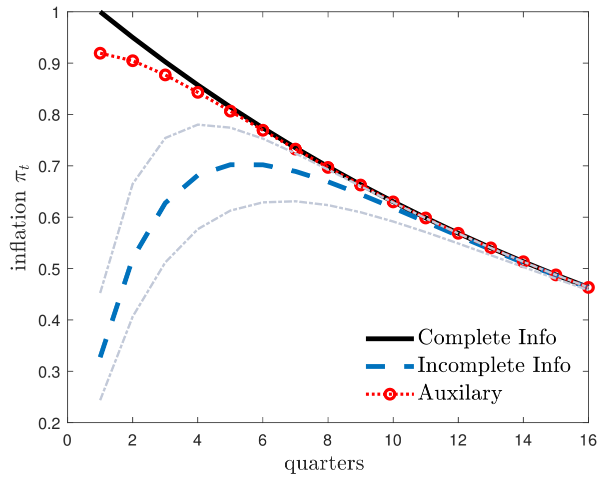

Figure 3: Response of Inflation to Higher Real Marginal Cost

K_{CG} equal to 1.193, with a standard deviation of 0.185. Translating the 95% confidence interval through the mapping developed in Section 5 yields the darker, thicker segment in Figure 2. This segment thus identifies the combinations of (ω_f, ω_b) that can be rationalized with a level of noise consistent with the expectation evidence in CG.

Clearly, only the third of the three estimates provided by Galí, Gertler, and Lopez-Salido (2005), that corresponding to the furthest right point in the figure, is noticeably away from this segment. This happens to be the estimate that these authors trust the least for independent, econometric, reasons. We conclude that, when the theory is disciplined by the evidence in CG, it generates distortions broadly in line with existing estimates of the Hybrid NKPC. More succinctly, the informational friction implicit in the expectations data may alone account for all the observed inertia in inflation.

6.4 A Decomposition

The quantitative implications of the theory are further illustrated in Figure 3. This figure compares the impulse response function of inflation under three scenarios. The solid line corresponds to frictionless benchmark. The dashed line corresponds to the frictional case, calibrated to the mean estimate of K_{CG} reported above. The circled dotted line is explained shortly.

As evident in the figure, the quantitative bite of the informational friction is significant: the impact effect on inflation is about 60% lower than its complete-information counterpart, and the peak of the inflation response is attained 5 quarters after impact rather than on impact. But what drives this quantitative bite? The lack of information about the real marginal cost (the PE component), or the beliefs about inflation (the GE component)?

The answer to this question is provided by the circled dotted line in Figure 3. This line represents a counterfactual that shuts down the effect of the informational friction on the expectations of the behavior of others (inflation) and isolates its effect on the the expectations of the fundamental (the real marginal cost). As evident in the figure, this counterfactual is very close to the complete-information benchmark and far away from the incomplete-information case. It follows that most of documented quantitative bite is due to the GE channel, or the anchoring of the expectations of inflation.[21]

6.5 Cross-Sectional Moments

Thus far we have disregarded the individual-level evidence of Bordalo et al. (2020) and Broer and Kohlhas (2019). For the reasons explained in Section 5, this evidence can be matched by letting agents be over- or under-confident, without influencing any of the preceding findings. This, however, does not mean that such evidence has no bite on the quantitative performance of the model. If we use the CG moment in combination with the individual-level counterpart estimated in the aforementioned papers, we can jointly identify $\hat{\sigma}$ and $\hat{\sigma}$, the perceived and the actual level of noise. We can then further test the model by looking at its predictions for other, non-targeted moments, such as the cross-sectional dispersion of the individual forecast errors or that of the individual forecast revisions.

We implement this test in Table 1. We continue to denote with K_{CG} the coefficient of regression (23), and we denote with $K_{\text{BGMS/BK}}$ the individual-level counterpart. We then consider three sets of estimates for these coefficients. The first corresponds to Coibion and Gorodnichenko (2015) and to the exercise conducted above. The second and the third sets are from Bordalo et al. (2020) and Broer and Kohlhas (2019), respectively.[22] For each set, we report the identified belief parameters, the implied degrees of myopia and anchoring, and the model's predictions about the aforementioned cross-sectional moments. We finally compare the latter to their empirical counterparts.

As explained in the legend of the table, we consider two possible normalizations of the cross-sectional moments. Some normalization is needed because the analysis so far has been silent about the scale of the fluctuations in inflation. In one, we normalize by the unconditional volatility of the quarter-to-quarter change in inflation. In the other, we normalize by the unconditional volatility of the level of inflation. We a priori prefer the first normalization, because our model is not supposed to capture low-frequency phenomena (e.g., great moderation) that may be "polluting" the second measure. But the model does a good job in both cases.

[21] The decomposition offered in Figure 3 mirrors the decomposition of PE and GE effects introduced in Section 3.1. See Appendix D.2 for the detailed construction.

[22] Though both papers confirm that the original CG findings that K_{CG} is positive, they disagree on the sign of K_{BGMS}. This reflects differences in the treatment of outliers and other implementation details.

Table 1: Moments on Average and Individual Inflation Forecasts

	K_{CG}	$K_{\text{BGMS/BK}}$	$\hat{\sigma}$	σ	ω_f	ω_b
CG	1.19	0.00	1.76	1.76	0.52	0.43
BGMS	1.41	0.18	2.04	1.61	0.48	0.46
BK	1.27	-0.19	1.86	2.61	0.51	0.44

	Forecast error dispersion				Forecast revision dispersion			
	data[1]	model[1]	data[2]	model[2]	data[1]	model[1]	data[2]	model[2]
CG	2.08	2.03	0.40	0.24	1.97	1.63	0.38	0.19
BGMS	2.08	1.80	0.40	0.20	1.97	1.32	0.38	0.14
BK	2.08	2.98	0.40	0.34	1.97	2.31	0.38	0.26

Note: The three rows correspond to different estimates for K_{CG}, the coefficient of regression (23), and $K_{\text{BGMS/BK}}$, the individual-level counterpart. In the first row, K_{CG} is taken form Panel B, Table 1 of Coibion and Gorodnichenko (2015), and $K_{\text{BGMS/BK}}$ is fixed to zero. In the second row, both K_{CG} and $K_{\text{BGMS/BK}}$ are taken from Table 3 of Bordalo et al. (2020). And in the third row, they are taken from Table 1 of Broer and Kohlhas (2019). The columns under *forecast error dispersion* correspond to the standard deviation of the cross-sectional forecast errors normalized by the standard deviation of either the quarter-to-quarter change in inflation (columns with superscript 1) or the level of inflation (with superscript 2). The columns under *forecast revision dispersion* correspond to the standard deviation of the cross-sectional forecast revisions with the same normalizations.

6.6 Food for Thought

We wrap up our application to inflation with a few additional insights about the possible determinants of the Hybrid NKPC implied by our analysis.

We start by studying the role of market concentration.[23] To this goal, we modify the micro-foundations as follows. There is now a continuum of markets, in each of which there is a finite number, $N \geq 2$, of competitors. We index the markets by $m \in [0,1]$ and the firms in a given market by $i \in \{1, ..., N\}$. We let consumers have nested-CES preferences, so that the demand faced by firm i in market m is given by

$$Y_{i,m,t} = \left(\frac{P_{i,m,t}}{P_{m,t}}\right)^{-\psi} \left(\frac{P_{m,t}}{P_t}\right)^{-\varepsilon} Y_t,$$

where $P_{i,m,t}$ is the price of that firm, $P_{m,t}$ is the price index of the market that firm operates in, P_t is the aggregate price level, Y_t is aggregate income, $\psi > 1$ is the within-market elasticity of substitution and $\epsilon \in (0, \psi)$ is the cross-market counterpart. We finally assume that each firm has complete information about its own market but incomplete information about the entire economy.[24]

[23] We thank a referee for suggesting this direction.

[24] The logic for the offered result requires only that information is more correlated within a market than across markets, or that firms face less higher-order uncertainty about their immediate links in the market network than about their remote links. The sharper assumption that firms face no higher-order uncertainty about their immediate links only simplifies the exposition.

Proposition 8. *In the economy described above, Corollary 2 continues to hold, modulo the following modification: both distortions decrease with market concentration (i.e., they increase with N).*

The intuition behind this result is that a higher degree of market concentration increases the strategic complementarity *within* markets and decreases it *across* markets. To the extent that firms know more about their own market than about the entire economy, this amounts to a lower bite of higher-order uncertainty, and therefore less myopia and less anchoring in the aggregate inflation dynamics.

This result links two empirical trends: the increase in market concentration (De Loecker, Eeckhout, and Unger, 2020; Autor et al., 2020) and the reduction in inflation persistence (Cogley, Primiceri, and Sargent, 2010; Fuhrer, 2010). Of course, this correlation does not establish causality. Still, the result illustrates how our analysis sheds new light on the possible determinants of inflation persistence.

We conclude with two additional ideas along these lines. The first one regards the conduct of monetary policy. Under the lens of our approach, a more hawkish monetary policy, such as that followed in the post-Volker era, is predicted to contribute to lower inflation persistence by reducing the effective degree of strategic complementarity in the firms' pricing decisions.

The second idea regards the economy's input-output structure. Rubbo (2020) has recently argued, in a setting abstracting from informational frictions, that changes in the input-output structure help explain the flattening of the NKPC. Our own analysis suggests that, in the presence of informational frictions, such changes may have also influenced the endogenous persistence in inflation, or the backward-looking component of the Hybrid NKPC.[25]

The exploration of these ideas is left for future work. But Section 8 paves the way for them by extending our tools to multi-variate systems and networks.

7 Application to Consumption and Bridge to HANK

Now we turn to the effects of incomplete information on aggregate demand. As already shown in Corollary 3, the Euler equation is modified as if there is additional discounting together with habit persistence. In this section, we illustrate the quantitative potential of this idea. We also build a bridge to the HANK literature by showing that the habit-like sluggishness generated by the informational friction is amplified when the agents with the highest MPC are also the ones with the most cyclical income (Patterson, 2019; Flynn, Patterson, and Sturm, 2019).

[25] La'O and Tahbaz-Salehi (2020) make a similar point as Rubbo (2020) in a setting where nominal rigidity originates in incomplete information, but abstract from forward-looking behavior and learning, which are the forces highlighted here.

7.1 A HANK-like Extension

We consider a perpetual-youth, overlapping-generations version of the New Keynesian model, along the lines of Piergallini (2007), Del Negro, Giannoni, and Patterson (2015), and Farhi and Werning (2019). As in those papers, finite horizons (mortality risk) serve as convenient proxies for liquidity constraints, self-control problems, and other micro-level frictions that help explain why most estimates of the MPC in microeconomic data are almost an order of magnitude larger than that predicted by the textbook infinite-horizon model. We take this basic insight a step further by letting heterogeneity in mortality risk capture heterogeneity in the MPC. We couple this with heterogeneity in cyclical exposure. And, crucially, we let information be incomplete.

There are n types, or groups, of consumers, indexed by $g \in \{1, ..., n\}$, with respective mass π_g. In each period, a consumer in group g remains alive with probability $\varpi_g \in (0,1]$; with the remaining probability, she dies and gets replaced by a new consumer of the same type. Consumers can trade actuarially fair annuities, so the return to saving, conditional on survival, is R_t/ϖ_g. This makes sure that the mortality risk does not distort intertemporal smoothing. Still, heterogeneity in ω_g matters because it maps to heterogeneity in MPCs. On top of that, different groups can have different exposures to the business cycle: the (log) income of group g is $y_{g,t} = \phi_g y_t$, where $\phi_g \geq 0$ is the elasticity of that group's income with respect to aggregate income and $\sum_g \pi_g \phi_g = 1$.

These assumptions allow us to study how the propagation mechanism under consideration, namely that related to incomplete information and higher-order beliefs, depends on heterogeneity in MPCs and business-cycle exposures. But they also open the door to a *separate* propagation mechanism: the dynamics of wealth inequality and the associated role of fiscal policy. To isolate the effects of interest, to nest the present application to the abstract analysis of Section 8, and to obtain a sharp theoretical result (Proposition 9 below), we neutralize the second mechanism by letting appropriate fiscal transfers undo any wealth inequality triggered by interest-rate shocks.[26]

As shown in Appendix E, the group-level spending can be expressed as follows:

$$c_{g,t} = m_g \phi_g \sum_{k=0}^{\infty} (1-m_g)^k \overline{\mathbb{E}}_t^g [c_{t+k}] - (1-m_g) \sum_{k=0}^{\infty} (1-m_g)^k \overline{\mathbb{E}}_t^g [r_{t+k}], \tag{26}$$

where $m_g \equiv 1 - \chi \varpi_g$, χ is the subjective discount rate, and $\overline{\mathbb{E}}_t^g$ is the average expectation. For each g, equation (26) follows from aggregating the consumption functions of the individuals within group g and replacing their income in terms of aggregate consumption. The collection of these equations across g recasts the demand block of the economy as a dynamic network among the various groups of consumers. This echoes Auclert, Rognlie, and Straub (2019), which develops similar network representations for more general HANK economies.

[26] An earlier draft had not clarified this assumption, without which the wealth distribution becomes a relevant state variable for the aggregate dynamics. We thank Dmitriy Sergeyev for pointing out this. See Appendix E for details.

Inspection of (26) reveals, first, that m_g identifies the MPC of group g and, second, that the strategic complementarity, or the Keynesian cross, depends on how the product $m_g \phi_g$ varies across groups, or whether a higher MPC is positively correlated with a higher business-cycle exposure. Patterson (2019) provides evidence of such a positive correlation and shows how it translates to a steeper Keynesian cross in a static, complete-information context. In the light of our insight of how the as-if distortions introduced by informational frictions depend on GE feedback mechanisms, one may expect such a positive correlation to translate also to more myopia and habit-like persistence in the aggregate consumption dynamics. We verify this intuition in part (iii) below, at least under the simplifying assumption of two groups.

Proposition 9 (HANK)**.** *(i) Under complete information, there exists a scalar $\varsigma > 0$ such that aggregate consumption obeys a textbook Euler condition of the following form:*

$$c_t = -\varsigma r_t + \mathbb{E}_t[c_{t+1}].$$

(ii) Under incomplete information, there exist scalars $\omega_f < 1$ and $\omega_b > 0$ such that aggregate consumption obeys a hybrid Euler condition of the form:

$$c_t = -\varsigma r_t + \omega_f \mathbb{E}_t[c_{t+1}] + \omega_b c_{t-1},$$

where the scalar $\varsigma > 0$ is the same as that under complete information and the scalars $\omega_f < 1$ and $\omega_b > 0$ are functions of σ and $(\pi_g, m_g, \phi_g)_{g \in \{1,\ldots,n\}}$.

(iii) Suppose there are two groups, with $m_1 > m_2$. An increase in ϕ_1, the business-cycle exposure of high-MPC group, maps to a lower ω_f and a higher ω_b, that is, more as-if myopia and anchoring in the aggregate dynamics.

Part (i) mirrors an irrelevance result from Werning (2015). With complete information, the DIS curve of our HANK economy is the same as a representative agent's Euler condition. There is neither extra discounting of the future nor habit-like persistence. Heterogeneity matters at most for ς, the elasticity of aggregate consumption with respect to the real interest rate.

Part (ii) extends Corollary 3 to heterogeneity in MPC and business-cycle exposure. Once again, incomplete information amounts to adding myopia and habit-like persistence in the DIS curve. But now heterogeneity interacts with information in shaping the magnitude of these distortions.

Part (iii) completes the picture by showing how exactly heterogeneity matters. An increase in the business-cycle exposure of the high-MPC group (and a corresponding reduction in the business-cycle exposure of the low-MPC group) translates to both more myopia and more habit-like persistence.

The basic logic behind this result was anticipated above. Its proof utilizes the techniques developed in Section 8. In the remainder of this section, we use a numerical example to illustrate our findings.

7.2 Numerical Example

Figure 4 compares four economies. The first one corresponds to the textbook, representative-agent benchmark. We refer to this benchmark as "Complete Information" in the figure. The second economy is a variant of the first one that adds habit persistence, of the type and magnitude found in the DSGE literature.[27] We refer to this economy as "Complete Info + Habit." The remaining two economies remove habit but add incomplete information. Both of them feature an average MPC equal to $\overline{m} = .30$, which is roughly consistent with the relevant evidence. The one referred to as "Incomplete Info" in the figure, abstracts from heterogeneity; this is the economy described in Corollary 3. The other one, which is referred to as "Incomplete Info + HANK" in the figure, adds heterogeneity: there are two groups of consumers, with $m_1 = .55$, $m_2 = .05$, $\phi_1 = 2$, and $\phi_2 = 0$.[28]

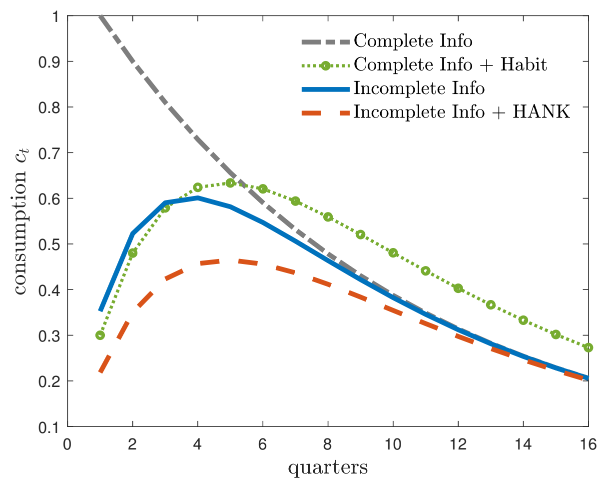

Figure 4: Response of Consumption to Lower Interest Rates

Let us first compare "Incomplete Info" to "Complete Info + Habit." This extends the lesson of the previous section from the inflation context to the consumption context: the informational friction alone generates a similar degree of sluggishness as that generated by habit persistence in the DSGE literature. Importantly, whereas the degree of habit assumed in that literature is far larger than that supported by micro-economic evidence (Havranek, Rusnak, and Sokolova, 2017), the informational friction assumed

[27] In particular, we assume external habit and specify the per-period utility as $\log(C_t - b\bar{C}_t)$, where C_t and $\bar{C}_t$ denote, respectively, own consumption and aggregate consumption. In equilibrium, $\bar{C}_t = C_t$ and the log-linearized Euler condition reduces to the following low of motion of consumption:

$$c_t = -\frac{1-b}{1+b} r_t + \frac{1}{1+b}\mathbb{E}_t[c_{t+1}] + \frac{b}{1+b} c_{t-1}.$$

We finally set $b = .7$, which is in the middle of the macro-level estimates reported in the meta-analysis by Havranek, Rusnak, and Sokolova (2017).

[28] For the incomplete-information economies, we target $K_{CG} = 0.9$. This is in the middle of the range of values Angeletos, Huo, and Sastry (2020) estimate when they repeat the CG regression on forecasts of unemployment, with the rationale being that unemployment is a proxy for the output gap in the model.

here is broadly consistent with survey evidence. This illustrates how our approach help merge the gap between the micro and macro estimates of habit.

Relatedly, if we consider an extension with transitory idiosyncratic income shocks along the lines of Appendix C, our economy can feature simultaneously two properties: a large and front-loaded response to such shocks at the micro level, in line with the relevant microeconomic evidence; and a dampened and sluggish response to monetary policy at the macro level, in line with the relevant macroeconomic evidence. By contrast, if there was true habit persistence in consumption of the kind and level assumed in the DSGE literature, the micro-level responses would also be dampened and sluggish, contradicting the relevant microeconomic evidence. This idea is pushed further, and is more carefully quantified, in a recent paper by Auclert, Rognlie, and Straub (2020).

Finally, let us inspect the economy "Incomplete Info + HANK." Needless to say, this economy is not meant to capture a realistic degree of heterogeneity: our two-group specification is only a gross approximation to the kind of heterogeneity captured in the quantitative HANK literature (e.g., Kaplan and Violante, 2014; Kaplan, Moll, and Violante, 2018)). Nevertheless, this economy helps illustrate how such heterogeneity, and in particular the kind of positive cross-sectional correlation between MPCs and income cyclicality documented in Patterson (2019), can reinforce both the habit-like sluggishness and the myopia-like dampening generated by incomplete information.

7.3 Informational Friction Plus Wealth Dynamics

In the preceding analysis we used appropriate fiscal transfers to make sure that the wealth distribution is not a state variable for the aggregate dynamics and to nest the exercise into the analysis of Section 8. We now shut down these transfers and study how the endogenous dynamics of wealth matter both in isolation and in combination with our mechanisms.

Consider first the case with complete information and suppose again that there are two groups, with only the high-MPC group being exposed to the business cycle ($\omega_1 < \omega_2$ and $\phi_1 > 0 = \phi_2$), and consider a negative innovation in η_t. This causes, in equilibrium, an expansion. But because only the first group's income is exposed to it, and because the income increase is less than permanent, this group will try to save some of this increase, while the second group has no such incentive. Along with the fact that the total saving of the two groups has to be zero, this explains why the first group responds to the shock by saving and accumulating wealth, whereas the second group responds by borrowing and accumulating debt. But since the first group has a larger MPC, the accumulation of wealth by this group helps increase aggregate spending in the future. This suggests that, even with complete information, the wealth dynamics add persistence to the response of aggregate demand to interest-rate shocks.

We verify this intuition in Figure 5 and proceed to show how this source of persistence extends to the

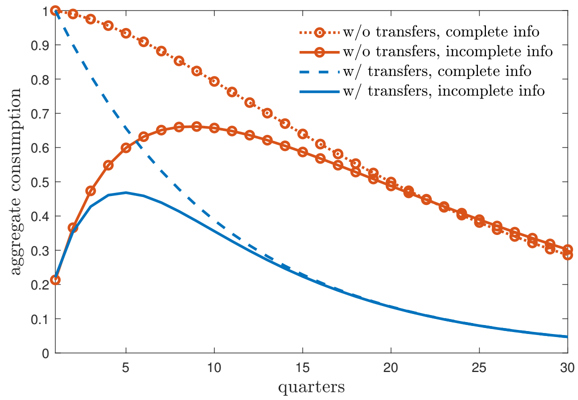

Figure 5: Shutting Down the Fiscal Transfers

case of incomplete information, without however upsetting, and indeed only reinforcing, our own message. This figure compares the response of consumption to a negative (expansionary) interest rate shock under four scenarios. Two of them replicate the complete-information and the incomplete-information HANK cases from Figure 4. The remaining two show how the results change when fiscal transfers are switched off and, equivalently, the aforementioned wealth channel is switched on. Regardless of the information structure, this channel adds persistence.[29] The effect of the informational friction, which is our own focal point, is qualitatively the same whether the wealth channel is present or not. Perhaps more interestingly, the two mechanisms reinforce each other, yielding a much more pronounced hump-shaped response than each mechanism alone.

8 Multivariate Systems, or Networks

We close the paper with the extension of our analytical results to multi-variate systems, or networks. We already made implicit use of this extension in our HANK application. Here, we fill in the details and develop tools that could aid analytical and quantitative evaluations of how informational frictions and network structures interact in a variety of applications.

The economy consists of n groups, each containing a continuum of agents. Groups are indexed by $g \in \{1, ..., n\}$, agents by (i, g) where $i \in [0, 1]$ is an agent's name and g her group affiliation (e.g., consumer or firm). The best response of agent i in group g is specified as follows:

$$a_{i,g,t} = \varphi_g \mathbb{E}_{i,g,t}[\xi_t] + \beta_g \mathbb{E}_{i,g,t}[a_{i,g,t+1}] + \sum_{j=0}^{n} \gamma_{gj} \mathbb{E}_{i,g,t}[a_{j,t+1}]. \tag{27}$$

The parameter φ_g captures the direct, contemporaneous exposure of an agent in group g to the exogenous shock, holding constant her expectations of both her own future actions and the actions of others.

[29] This channel also adds amplification. To focus on the persistence effects, in the figure we renormalize the magnitude of the shock as we change the fiscal rule so that the complete-information response of consumption on impact remains 1.

The parameter $\{\beta_g\}$ captures the additional, forward-looking, PE effect that obtains because of the consideration of own future actions. Finally, the parameter $\{\gamma_{g,j}\}$ captures the dependence of the optimal action of an agent in group g to her expectation of the average action of group j. This allows for rich strategic of GE interactions both within groups (when $j = g$) and across groups (when $j \neq g$).[30]

Turning now to the information structure, this is specified as a collection of private Gaussian signals, one per agent and per period. The period-t signal received by agent i in group t is given by

$$x_{i,g,t} = \xi_t + u_{i,g,t}, \quad u_{i,g,t} \sim \mathcal{N}(0, \sigma_g^2). \tag{28}$$

where $\sigma_g \geq 0$ parameterizes the noise of group g. Notice that, by allowing σ_g to differ across g, we can accommodate information heterogeneity in addition to payoff and strategic heterogeneity. For instance, firms could be more informed than consumers on average, and "sophisticated" consumers could be more informed than "unsophisticated" ones.

Let $\boldsymbol{a}_t = (a_{g,t})$ be a column vector collecting the aggregate actions of all the groups (e.g., the vector of aggregate consumption and aggregate inflation). Let $\boldsymbol{\varphi} = (\varphi_g)$ be a column vector containing the value of φ_g across the groups. Let $\boldsymbol{\beta} = \mathbf{diag}\{\beta_g\}$ be a $n \times n$ diagonal matrix whose off-diagonal elements are zero and whose diagonal elements are the values of β_g across groups. Finally, let $\boldsymbol{\gamma} = (\gamma_{gk})$ be an $n \times n$ matrix collecting the interaction parameters, γ_{gj}, and let $\boldsymbol{\delta} \equiv \boldsymbol{\beta} + \boldsymbol{\gamma}$. Similarly to Section 2, we impose that $\beta_g \in (0,1)$ and the spectral radius of $(\mathbf{I} - \boldsymbol{\beta})^{-1}\boldsymbol{\gamma}$ is less than 1. The following extensions of Propositions 2 and 3 are then possible.

Proposition 10 (Solution)**.** *There exists a unique equilibrium, and the aggregate outcome $a_{g,t}$ of each group g is given by*

$$a_{g,t} = \sum_{j=1}^{n} \psi_{g,j} \left\{ \frac{1 - \frac{\vartheta_j}{\rho}}{1 - \vartheta_j L} \xi_t \right\}, \tag{29}$$

where $\{\psi_{g,j}\}$ are fixed scalars, characterized in Appendix A, and $\{\vartheta_g\}$ are the inverse of the outside roots of the following polynomial:

$$C(z) = \det\left((\boldsymbol{\delta} - \boldsymbol{\gamma} - \mathbf{I}z)\, \mathbf{diag}\left\{ z^2 - \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2} \right) z + 1 \right\} - z\, \mathbf{diag}\left\{ \frac{1}{\rho\sigma_g^2} \right\} \boldsymbol{\gamma} \right). \tag{30}$$

Proposition 11 (Observational Equivalence)**.** *There exist matrices $\boldsymbol{\omega}_f$ and $\boldsymbol{\omega}_b$ such that the incomplete-information economy is observationally equivalent to the following complete-information economy:*

$$\boldsymbol{a}_t = \boldsymbol{\varphi}\xi_t + \boldsymbol{\omega}_f \boldsymbol{\delta} \mathbb{E}_t[\boldsymbol{a}_{t+1}] + \boldsymbol{\omega}_b \boldsymbol{a}_{t-1}. \tag{31}$$

[30] Like our baseline framework, the extension considered here rules out the dependence of an agent's best response on the *concurrent* choices of others. This, however, is without serious loss of generality for two reasons. First, in all applications of interest, this dependence vanishes as the length of the time period goes to zero. Second, if we incorporate a general form of such dependence by adding the term $\sum_j \alpha_{g,i} \mathbb{E}_{i,j,t}[a_{i,j,t}]$ in equation (27), the results stated below, namely Propositions 10 and 11, continue to hold, modulo a minor adjustment in the cubic that appears in condition (29).

One subtlety with representation (31) is that it is not unique: there are multiple values of the matrices $\boldsymbol{\omega_f}$ and $\boldsymbol{\omega_b}$ that replicate the incomplete-information equilibrium. Intuitively, it is possible to make agents myopic vis-a-vis the future by letting them discount enough either only their own group's future actions, or the future actions of other groups too.[31] This complicates the interpretation and the comparative statics of the provided representation but is of little substantial consequence: although the representation in terms of condition (31) is not unique, the equilibrium itself is determinate, and so are its observable properties, which can be directly obtained from Proposition 10.

Proposition 10 is indeed quite telling. It shows that the equilibrium outcome can now be expressed as a linear combination of n terms, each of which is an AR(2) process that has a similar structure as in our baseline analysis. The one root of these processes is the same across g and is given, naturally, by that of the fundamental. The other root, denoted above by ϑ_g, encodes how the information friction faced by group g interacts with the network structure of the economy.

In the knife-edge case in which $\boldsymbol{\gamma}$ is diagonal, meaning that the behavior of each group is independent of that of other groups, each ϑ_g is pinned down by the characteristics of group g alone and the outcome of that group is given by the corresponding AR(2) process alone ($\psi_{g,j} = 0$ for $j \neq g$). For generic $\boldsymbol{\gamma}$, instead, each ϑ_g depends on the entire $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ matrices, that is, on all the PE and GE parameters, as well as on all the information parameters. Furthermore, the outcome of a group depends on all the n different AR(2) processes.

To illustate how the network structure matters, let $\boldsymbol{\beta} = 0$ and $\sigma_g = \sigma$ for all g. In this case, we show in Appendix A that the polynomial given in condition (30) reduces to the product of n quadratics, one for each ϑ_g. Furthermore, each ϑ_g is determined in the same manner as in our baseline analysis, namely as the reciprocal of the largest solution of cubic (17), with the g-th eigenvalue of the matrix $\boldsymbol{\gamma}$ in place of the scalar γ. Because the eigenvalues of $\boldsymbol{\gamma}$ encode the GE feedback both within and across groups, we have that an increase in either kind of feedback maps to a higher ϑ_g and, thereby, to both less amplitude and more volatility. The essence of our baseline analysis is thus fully preserved.

Finally, note that the results presented here not only offer a robustness of our main insights to multivariate systems and networks, but also a straightforward numerical algorithm: one only needs to solve the polynomial in condition (30).

9 Conclusion

We developed a toolbox for analyzing and quantifying the equilibrium effects of informational frictions and of the associated higher-order uncertainty. We represented these effects as the combination of two

[31] Indeed, both of the following two choices are possible: let $\boldsymbol{\omega}_f$ have unit off-diagonal elements, meaning that a distortion is applied only to expectations of own-group future outcomes; or let the elements of each row of $\boldsymbol{\omega}_f$ be the same, meaning that the same distortion is applied to all expectations. If one of these choices is made, there is no residual indeterminacy.

behavioral distortions: a form of myopia, or extra discounting of the future; and a form of habit, or anchoring of current behavior to past behavior. We further showed how these as-if distortions increase with the strength of the underlying strategic interaction or GE feedback, and how they can be disciplined with available evidence on expectations. And we used these results to argue that the friction implicit in survey evidence of expectations is large enough to generate a comparable amount of sluggishness in the dynamics of inflation and aggregate spending as that captured in the DGSE literature with more ad hoc modeling devices.

While connecting the theory to the available evidence on expectations, we clarified *which* such evidence is best suited for the purpose of quantifying the distortions of interest: it is evidence on *average* forecasts, such as that provided in Coibion and Gorodnichenko (2015), as opposed to evidence on *individual* forecasts, such as that provided in Bordalo et al. (2020) and Broer and Kohlhas (2019). Left outside this paper was a more comprehensive investigation of the lessons contained in surveys of expectations for macroeconomic theory.

We undertake this task in a follow-up paper (Angeletos, Huo, and Sastry, 2020). There, we use a variety of existing evidence along with new evidence of our own to argue that, among a large set of candidate theories, the one that best accounts for the joint dynamics of inflation, aggregate spending and forecasts thereof in the US is a theory that blends two frictions: incomplete information or rational inattention, as in the present paper and the literature we have built on; and over-extrapolation, as in Greenwood and Shleifer (2014) and Gennaioli, Ma, and Shleifer (2015). This points in the opposite direction than cognitive discounting and level-k thinking, two close cousins of under-extrapolation, but leaves room for the kinds of myopia and anchoring accommodated via our approach.

Another element of our contribution was to extend our tools to multi-variate systems and networks. We illustrated the use of these extended tools within a HANK economy. Other possible applications include production networks, whether in the context of the NKPC (La'O and Tahbaz-Salehi, 2020; Rubbo, 2020) or in the context of the RBC framework (Acemoglu et al., 2012; Baqaee and Farhi, 2019; Nimark, Chahrour, and Pitschner, 2019), as well as dynamic extensions of the more abstract incomplete-information networks studied in Bergemann, Heumann, and Morris (2017).

References

Abel, Andrew B. and Olivier J. Blanchard. 1983. "An Intertemporal Model of Saving and Investment." *Econometrica* 51 (3):675–692.

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations." *Econometrica* 80 (5):1977–2016.

Allen, Franklin, Stephen Morris, and Hyun Song Shin. 2006. "Beauty Contests and Iterated Expectations in Asset Markets." *Review of financial Studies* 19 (3):719–752.

Alvarez, Fernando and Francesco Lippi. 2014. "Price Setting with Menu Cost for Multiproduct Firms." *Econometrica* 82 (1):89–135.

Angeletos, George-Marios, Fabrice Collard, and Harris Dellas. 2019. "Business Cycle Anatomy." *American Economic Review* conditionally accepted.

Angeletos, George-Marios, Zhen Huo, and Karthik Sastry. 2020. "Imperfect Macroeconomic Expectations: Evidence and Theory." *prepared for the 35th NBER Macroeconomics Annual* .

Angeletos, George-Marios and Chen Lian. 2018. "Forward Guidance without Common Knowledge." *American Economic Review* 108 (9):2477–2512.

Angeletos, George-Marios and Alessandro Pavan. 2007. "Efficient Use of Information and Social Value of Information." *Econometrica* 75 (4):1103–1142.

Auclert, Adrien, Matthew Rognlie, and Ludwig Straub. 2019. "The Intertemporal Keynesian Cross." *mimeo* .

———. 2020. "Micro Jumps, Macro Humps: monetary policy and business cycles in an estimated HANK model." *mimeo* .

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The fall of the labor share and the rise of superstar firms." *The Quarterly Journal of Economics* 135 (2):645–709.

Bachmann, Rüdiger, Ricardo J. Caballero, and Eduardo M. R. A. Engel. 2013. "Aggregate Implications of Lumpy Investment: New Evidence and a DSGE Model." *American Economic Journal: Macroeconomics* 5 (4):29–67.

Baqaee, David Rezza and Emmanuel Farhi. 2019. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica* 87 (4):1155–1203.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris. 2017. "Information and Interaction." *Cowles Foundation Discussion Paper* (2088).

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2020. "Over-reaction in Macroeconomic Expectations." *The American Economic Review* forthcoming.

Broer, Tobias and Alexandre Kohlhas. 2019. "Forecaster (Mis-)Behavior." *IIES miméo* .

Caballero, Ricardo J. and Eduardo M. R. A. Engel. 1999. “Explaining Investment Dynamics in U.S. Manufacturing: A Generalized (S, s) Approach.” *Econometrica* 67 (4):783–826.

Carroll, Christopher D., Edmund Crawley, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. 2020. “Sticky Expectations and Consumption Dynamics.” *American Economic Journal: Macroeconomics* forthcoming.

Carvalho, Carlos, Stefano Eusepi, Emanuel Moench, and Bruce Preston. 2017. “Anchored Inflation Expectations.” *mimeo* .

Choi, Yongok, Giacomo Rondina, and Todd B Walker. 2020. “Information Aggregation Bias and SamuelsonâĂŹs Dictum.” *mimeo* .

Christiano, Lawrence J, Martin Eichenbaum, and Charles L Evans. 2005. “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy.” *Journal of Political Economy* 113 (1):1–45.

Cogley, Timothy, Giorgio E Primiceri, and Thomas J Sargent. 2010. “Inflation-gap persistence in the US.” *American Economic Journal: Macroeconomics* 2 (1):43–69.

Coibion, Olivier and Yuriy Gorodnichenko. 2012. “What Can Survey Forecasts Tell Us about Information Rigidities?” *Journal of Political Economy* 120 (1):116–159.

———. 2015. “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts.” *American Economic Review* 105 (8):2644–78.

De Loecker, Jan, Jan Eeckhout, and Gabriel Unger. 2020. “The rise of market power and the macroeconomic implications.” *The Quarterly Journal of Economics* 135 (2):561–644.

Del Negro, Marco, Marc P Giannoni, and Christina Patterson. 2015. “The Forward Guidance Puzzle.” *FRB of New York mimeo* .

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

Farhi, Emmanuel and Iván Werning. 2019. “Monetary Policy, Bounded Rationality, and Incomplete Markets.” *American Economic Review* 109 (11):3887–3928.

Flynn, Joel, Christina Patterson, and John Sturm. 2019. “Shock Propagation and the Fiscal Multiplier: The Role of Heterogeneity.” *MIT and Northwestern University mimeo* .

Fuhrer, Jeffrey C. 2010. “Inflation persistence.” In *Handbook of monetary economics*, vol. 3. Elsevier, 423–486.

Gabaix, Xavier. 2020. "A Behavioral New Keynesian Model." *American Economic Review* 110 (8):2271–2327.

Galí, Jordi. 2008. *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications Second edition.* Princeton University Press.

Galí, Jordi and Mark Gertler. 1999. "Inflation dynamics: A structural econometric analysis." *Journal of Monetary Economics* 44 (2):195–222.

Galí, Jordi, Mark Gertler, and J David Lopez-Salido. 2005. "Robustness of the estimates of the hybrid New Keynesian Phillips curve." *Journal of Monetary Economics* 52 (6):1107–1118.

Garcıa-Schmidt, Mariana and Michael Woodford. 2019. "Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis." *American Economic Review* 109 (1):86–120.

Gennaioli, Nicola, Yueran Ma, and Andrei Shleifer. 2015. "Expectations and investment." In *NBER Macroeconomics Annual 2015, Volume 30*. University of Chicago Press.

Golosov, Mikhail and Robert E Lucas Jr. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy* 115 (2):171–199.

Golub, Ben and Stephen Morris. 2019. "Expectations, Networks and Conventions." *mimeo* .

Greenwood, Robin and Andrei Shleifer. 2014. "Expectations of returns and expected returns." *Review of Financial Studies* 27 (3):714–746.

Havranek, Tomas, Marek Rusnak, and Anna Sokolova. 2017. "Habit Formation in Consumption: A Meta-Analysis." *European Economic Review* 95:142–167.

Hayashi, Fumio. 1982. "Tobin's Marginal q and Average q: A Neoclassical Interpretation." *Econometrica* 50 (1):213–224.

Huo, Zhen and Marcelo Zouain Pedroni. 2020. "A Single-Judge Solution to Beauty Contests." *American Economic Review* 110 (2):526–68.

Huo, Zhen and Naoki Takayama. 2018. "Rational Expectations Models with Higher Order Beliefs." miméo, Yale University.

Iovino, Luigi and Dmitriy Sergeyev. 2017. "Quantitative Easing without Rational Expectations." *Work in progress* .

Jung, Jeeman and Robert J. Shiller. 2005. "Samuelson's Dictum and the Stock Market." *Economic Inquiry* 43 (2):221–228.

Kaplan, Greg, Benjamin Moll, and Giovanni L. Violante. 2018. "Monetary Policy According to HANK." *American Economic Review* 108 (3):697–743.

Kaplan, Greg and Giovanni L. Violante. 2014. "A Model of the Consumption Response to Fiscal Stimulus Payments." *Econometrica* 82 (4):1199–1239.

Kasa, Kenneth, Todd B. Walker, and Charles H. Whiteman. 2014. "Heterogeneous Beliefs and Tests of Present Value Models." *The Review of Economic Studies* 81 (3):1137–1163.

Khan, Aubhik and Julia K. Thomas. 2008. "Idiosyncratic Shocks and the Role of Nonconvexities in Plant and Aggregate Investment Dynamics." *Econometrica* 76 (2):395–436.

Kiley, Michael T. 2007. "A Quantitative Comparison of Sticky-Price and Sticky-Information Models of Price Setting." *Journal of Money, Credit and Banking* 39 (s1):101–125.

Kohlhas, Alexandre and Ansgar Walther. 2019. "Asymmetric Attention." *IIES miméo* .

La'O, Jennifer and Alireza Tahbaz-Salehi. 2020. "Optimal Monetary Policy in Production Networks." *mimeo* .

Lucas, Robert E. Jr. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4 (2):103–124.

———. 1976. "Econometric Policy Evaluation: A Critique." *Carnegie-Rochester Conference Series on Public Policy* 1:19 – 46.

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

———. 2015. "Business Cycle Dynamics under Rational Inattention." *Review of Economic Studies* 82 (4):1502–1532.

Malmendier, Ulrike and Stefan Nagel. 2016. "Learning from inflation experiences." *The Quarterly Journal of Economics* 131 (1):53–87.

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

Matejka, Filip. 2016. "Rationally Inattentive Seller: Sales and Discrete Pricing." *Review of Economic Studies* 83 (3):1125–1155.

Mavroeidis, Sophocles, Mikkel Plagborg-Møller, and James H Stock. 2014. "Empirical evidence on inflation expectations in the New Keynesian Phillips Curve." *Journal of Economic Literature* 52 (1):124–88.

Melosi, Leonardo. 2016. "Signalling effects of monetary policy." *The Review of Economic Studies* 84 (2):853–884.

Midrigan, Virgiliu. 2011. "Menu Costs, Multiproduct Firms, and Aggregate Fluctuations." *Econometrica* 79 (4):1139–1180.

Morris, Stephen and Hyun Song Shin. 1998. "Unique Equilibrium in a Model of Self-fulfilling Currency Attacks." *American Economic Review* :587–597.

———. 2002. "Social Value of Public Information." *American Economic Review* 92 (5):1521–1534.

———. 2006. "Inertia of Forward-looking Expectations." *The American Economic Review* :152–157.

Nakamura, Emi and Jón Steinsson. 2013. "Price Rigidity: Microeconomic Evidence and Macroeconomic Implications." *Annual Review of Economics* 5 (1):133–163.

Nimark, Kristoffer. 2008. "Dynamic Pricing and Imperfect Common Knowledge." *Journal of Monetary Economics* 55 (2):365–382.

———. 2017. "Dynamic Higher Order Expectations." *Cornell Univeristy mimeo* .

Nimark, Kristoffer, Ryan Chahrour, and Stefan Pitschner. 2019. "Sectoral Media Focus and Aggregate Fluctuations." *mimeo* .

Patterson, Christina. 2019. "The Matching Multiplier and the Amplification of Recessions." *Northwestern University mimeo* .

Piergallini, Alessandro. 2007. "Real Balance Effects and Monetary Policy." *Economic Inquiry* 44 (3):497–511.

Reis, Ricardo. 2006. "Inattentive producers." *The Review of Economic Studies* 73 (3):793–821.

Romer, Christina D and David H Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American Economic Review* 94 (4):1055–1084.

Rubbo, Elissa. 2020. "Networks, Phillips Curves, and Monetary Policy." *mimeo* .

Sargent, Thomas J. 1993. *Bounded rationality in macroeconomics.* Oxford University Press.

Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3):665–690.

Singleton, Kenneth J. 1987. "Asset Prices in a Time-series Model with Disparately Informed, Competitive Traders." In *New Approaches to Monetary Economics*, edited by William A. Barnett and Kenneth J. Singleton. Cambridge University Press, 249–272.

Smets, Frank and Rafael Wouters. 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach." *American Economic Review* 97 (3):586–606.

Tirole, Jean. 2015. "Cognitive Games and Cognitive Traps." *Toulouse School of Economicss mimeo* .

Vives, Xavier and Liyan Yang. 2017. "Costly Interpretation of Asset Prices." mimeo, IESE/University of Toronto.

Werning, Iván. 2015. "Incomplete Markets and Aggregate Demand." *NBER Working Paper No. 21448* .

Whittle, Peter. 1963. "Prediction and Regulation by Linear Least-Square Methods." .

Woodford, Michael. 2003. "Imperfect Common Knowledge and the Effects of Monetary Policy." *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps* .

Zorn, Peter. 2018. "Investment under Rational Inattention: Evidence from US Sectoral Data." miméo, University of Munich.

ONLINE APPENDIX

The materials in this online appendix are organized as follows: Section A contains the proofs of propositions in the main text. The next two sections extend the main theoretical results in two different environments; Section B adds public signals and Section C introduces idiosyncratic fundamentals. Section D contains various results that complement the analysis of inflation with incomplete information in the main text. Section E contains the model details in the HANK application with incomplete information. Section F and Section G apply our observational equivalence result in the contexts of investment and asset prices, respectively. Section H generalizes the main insights in an environment with more flexible fundamental and signal processes. Section I shows how the observational equivalence result is modified when allowing the fundamental to be driven by multiple shocks. Section J contains proofs for additional propositions in this appendix.

A Proofs of Propositions in Main Text

Proof of Proposition 1

The proof follows from the main text.

Proof of Proposition 2

As a preliminary step, we look for the fundamental representation of the signals. Define $\tau_\eta = \sigma_\eta^{-2}$ and $\tau_u = \sigma^{-2}$ as the reciprocals of the variances of, respectively, the innovation in the fundamental and the noise in the signal. (In the main text, we have normalized $\sigma_\eta = 1$.) The signal process can be rewritten as

$$x_{i,t} = \mathbf{M}(L)\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,t}\end{bmatrix}, \qquad \text{with} \qquad \mathbf{M}(L) = \begin{bmatrix}\tau_\eta^{-\frac{1}{2}}\frac{1}{1-\rho L} & \tau_u^{-\frac{1}{2}}\end{bmatrix}.$$

Let $B(L)$ denote the fundamental representation of the signal process. By definition, $B(L)$ needs to be an invertible process that satisfies the following requirement

$$B(L)B(L^{-1}) = \mathbf{M}(L)\mathbf{M}'(L^{-1}) = \frac{\tau_\eta^{-1} + \tau_u^{-1}(1-\rho L)(L-\rho)}{(1-\rho L)(L-\rho)}. \tag{32}$$

This condition implies that

$$B(L) = \tau_u^{-\frac{1}{2}}\sqrt{\frac{\rho}{\lambda}}\frac{1-\lambda L}{1-\rho L},$$

where λ is the inside root of the numerator in the last term of equation (32)

$$\lambda = \frac{1}{2}\left[\rho + \frac{1}{\rho}\left(1+\frac{\tau_u}{\tau_\eta}\right) - \sqrt{\left(\rho + \frac{1}{\rho}\left(1+\frac{\tau_u}{\tau_\eta}\right)\right)^2 - 4}\right]. \tag{33}$$

The forecast of a random variable

$$f_t = \mathbf{A}(L)\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,t}\end{bmatrix}$$

can be obtained by using the Wiener-Hopf prediction formula:[32]

$$\mathbb{E}_{i,t}[f_t] = \left[\mathbf{A}(L)\mathbf{M}'(L^{-1})B(L^{-1})^{-1}\right]_+ B(L)^{-1}x_{i,t}.$$

Now we proceed to solve the equilibrium. Denote agents' equilibrium policy function as

$$a_{i,t} = h(L)x_{i,t}$$

for some lag polynomial $h(L)$. The aggregate outcome can then be expressed as follows:

$$a_t = h(L)\xi_t = \frac{h(L)}{1-\rho L}\eta_t.$$

In the sequel, we verify that the above guess is correct and characterize $h(L)$.

Consider the forecast of the fundamental. Note that

$$\xi_t = \begin{bmatrix}\tau_\eta^{-\frac{1}{2}}\frac{1}{1-\rho L} & 0\end{bmatrix}\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,t}\end{bmatrix},$$

from which it follows that

$$\mathbb{E}_{i,t}[\xi_t] = G_1(L)x_{i,t}, \qquad G_1(L) \equiv \frac{\lambda}{\rho}\frac{\tau_u}{\tau_\eta}\frac{1}{1-\rho\lambda}\frac{1}{1-\lambda L}.$$

Consider the forecast of the future own and average actions. Using the guess that $a_{it+1} = h(L)x_{i,t+1}$ and $a_{t+1} = h(L)\xi_{t+1}$, we have

$$a_{t+1} = \begin{bmatrix}\tau_\eta^{-\frac{1}{2}}\frac{h(L)}{L(1-\rho L)} & 0\end{bmatrix}\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,t}\end{bmatrix}, \quad a_{i,t+1} - a_{t+1} = \begin{bmatrix}0 & \tau_u^{-\frac{1}{2}}h(L)\end{bmatrix}\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,t}\end{bmatrix},$$

and the forecasts are

$$\mathbb{E}_{i,t}\left[a_{t+1}\right] = G_2(L)x_{i,t}, \qquad G_2(L) \equiv \frac{\lambda}{\rho}\frac{\tau_u}{\tau_\eta}\left(\frac{h(L)}{(1-\lambda L)(L-\lambda)} - \frac{h(\lambda)(1-\rho L)}{(1-\rho\lambda)(L-\lambda)(1-\lambda L)}\right),$$

$$\mathbb{E}_{i,t}\left[a_{i,t+1} - a_{t+1}\right] = G_3(L)x_{i,t}, \qquad G_3(L) \equiv \frac{\lambda}{\rho}\left(\frac{h(L)(L-\rho)}{L(L-\lambda)} - \frac{h(\lambda)(\lambda-\rho)}{\lambda(L-\lambda)} - \frac{\rho}{\lambda}\frac{h(0)}{L}\right)\frac{1-\rho L}{1-\lambda L}$$

Now, turn to the fixed point problem that characterizes the equilibrium:

$$a_{i,t} = \mathbb{E}_{i,t}[\varphi\xi_t + \beta a_{i,t+1} + \gamma a_{t+1}]$$

Using our guess, we can replace the left-hand side with $h(L)x_{i,t}$. Using the results derived above, on the other hand, we can replace the right-hand side with $\left[G_1(L) + (\beta+\gamma)G_2(L) + \beta G_3(L)\right]x_{i,t}$. It follows that

[32] See Whittle (1963) for more details about Wiener-Hopf prediction formula.

our guess is correct if and only if

$$h(L) = G_1(L) + (\beta+\gamma)G_2(L) + \beta G_3(L)$$

Equivalently, we need to find an analytic function $h(z)$ that solves

$$\begin{aligned} h(z) = {} & \varphi\frac{\lambda}{\rho}\frac{\tau_u}{\tau_\eta}\frac{1}{1-\rho\lambda}\frac{1}{1-\lambda z}+ \\ & +(\beta+\gamma)\frac{\lambda}{\rho}\frac{\tau_u}{\tau_\eta}\left(\frac{h(z)}{(1-\lambda z)(z-\lambda)}-\frac{h(\lambda)(1-\rho z)}{(1-\rho\lambda)(z-\lambda)(1-\lambda z)}\right) \\ & +\beta\frac{\lambda}{\rho}\left(\frac{h(z)(z-\rho)}{z(z-\lambda)}-\frac{h(\lambda)(\lambda-\rho)}{\lambda(z-\lambda)}-\frac{\rho}{\lambda}\frac{h(0)}{z}\right)\frac{1-\rho z}{1-\lambda z}, \end{aligned}$$

which can be transformed as

$$C(z)h(z) = d(z; h(\lambda), h(0))$$

where

$$\begin{aligned} C(z) &\equiv z(1-\lambda z)(z-\lambda) - \frac{\lambda}{\rho}\left\{\beta(z-\rho)(1-\rho z) + (\beta+\gamma)\frac{\tau_u}{\tau_\eta}z\right\} \\ d(z; h(\lambda), h(0)) &\equiv \varphi\frac{\lambda}{\rho}\frac{\tau_u}{\tau_\eta}\frac{1}{1-\rho\lambda}z(z-\lambda) - \frac{1}{\rho}\left(\frac{\tau_u}{\tau_\eta}\frac{\lambda(\beta+\gamma)}{1-\rho\lambda}+\beta(\lambda-\rho)\right)z(1-\rho z)h(\lambda) \\ & -\beta(z-\lambda)(1-\rho z)h(0) \end{aligned}$$

Note that $C(z)$ is a cubic equation and therefore contains with three roots. We will verify later that there are two inside roots and one outside root. To make sure that $h(z)$ is an analytic function, we choose $h(0)$ and $h(\lambda)$ so that the two roots of $d(z; h(\lambda), h(0))$ are the same as the two inside roots of $C(z)$. This pins down the constants $\{h(0), h(\lambda)\}$, and therefore the policy function $h(L)$

$$h(L) = \left(1-\frac{\vartheta}{\rho}\right)\frac{\varphi}{1-\rho\delta}\frac{1}{1-\vartheta L},$$

where ϑ^{-1} is the root of $C(z)$ outside the unit circle.

Now we verify that $C(z)$ has two inside roots and one outside root. $C(z)$ can be rewritten as

$$C(z) = \lambda\left\{-z^3 + \left(\rho+\frac{1}{\rho}+\frac{1}{\rho}\frac{\tau_u}{\tau_\eta}+\beta\right)z^2 - \left(1+\beta\left(\rho+\frac{1}{\rho}\right)+\frac{\beta+\gamma}{\rho}\frac{\tau_u}{\tau_\eta}\right)z+\beta\right\}.$$

With the assumption that $\beta > 0$, $\gamma > 0$, and $\beta+\gamma < 1$, it is straightforward to verify that the following properties hold:

$$\begin{aligned} C(0) &= \beta > 0 \\ C(\lambda) &= -\lambda\gamma\frac{1}{\rho}\frac{\tau_u}{\tau_\eta} < 0 \\ C(1) &= \frac{\tau_u(1-\beta-\gamma)}{\tau_\eta\rho} + (1-\beta)\left(\frac{1}{\rho}+\rho-2\right) > 0 \end{aligned}$$

Therefore, the three roots are all real, two of them are between 0 and 1, and the third one ϑ^{-1} is larger than 1.

Finally, to show that ϑ is less than ρ, it is sufficient to show that

$$C\left(\frac{1}{\rho}\right) = \frac{\tau_u(1-\rho\beta-\rho\gamma)}{\tau_\eta \rho^3} > 0.$$

Since $C(\vartheta^{-1}) = 0$, it has to be that ϑ^{-1} is larger than ρ^{-1}, or $\vartheta < \rho$.

Proof of Proposition 3

The equilibrium outcome in the hybrid economy is given by the following AR(2) process:

$$a_t = \frac{\zeta_0}{1-\zeta_1 L}\xi_t,$$

where

$$\zeta_1 = \frac{1}{2\omega_f \delta}\left(1-\sqrt{1-4\delta\omega_f\omega_b}\right) \quad \text{and} \quad \zeta_0 = \frac{\varphi\zeta_1}{\omega_b - \rho\omega_f\delta\zeta_1}, \tag{34}$$

and $\delta \equiv \beta+\gamma$. The solution to the incomplete-information economy is

$$a_t = \left(1-\frac{\vartheta}{\rho}\right)\frac{\varphi}{1-\rho\delta}\frac{1}{1-\vartheta L}\xi_t.$$

To match the hybrid model, we need

$$\zeta_1 = \vartheta \quad \text{and} \quad \zeta_0 = \left(1-\frac{\vartheta}{\rho}\right)\frac{\varphi}{1-\rho\delta}. \tag{35}$$

Combining (34) and (35), and solving for the coefficients of ω_f and ω_b, we infer that the two economies generate the same dynamics if and only if the following two conditions hold:

$$\omega_f = \frac{\delta\rho^2-\vartheta}{\delta(\rho^2-\vartheta^2)}, \tag{36}$$

$$\omega_b = \frac{\vartheta(1-\delta\vartheta)\rho^2}{\rho^2-\vartheta^2}. \tag{37}$$

Since $\delta \equiv \beta+\gamma$ and since ϑ is a function of the primitive parameters $(\sigma, \rho, \beta, \gamma)$, the above two conditions give the coefficients ω_f and ω_b as as functions of the primitive parameters, too.

It is immediate to check that $\omega_f < 1$ and $\omega_b > 0$ if $\vartheta \in (0,\rho)$, which in turn is necessarily true for any $\sigma > 0$; and that $\omega_f = 1$ and $\omega_b = 0$ if $\vartheta = \rho$, which in turn is the case if and only if $\sigma = 0$.

Proof of Propositions 4 and 5

To prove the comparative statics, we first show that ω_f is decreasing in ϑ and ω_b is increasing in ϑ. This can be verified as follows

$$\frac{\partial\omega_f}{\partial\vartheta}=\frac{-\delta(\rho^2+\vartheta^2)+2\delta^2\rho^2\vartheta}{(\delta(\rho^2-\vartheta^2))^2}<\frac{-\delta(\rho^2+\vartheta)+2\delta\rho\vartheta}{(\delta(\rho^2-\vartheta^2))^2}=\frac{-\delta(\rho-\vartheta)^2}{(\delta(\rho^2-\vartheta^2))^2}<0,$$
$$\frac{\partial\omega_b}{\partial\vartheta}=\frac{\rho^2(\rho^2+\vartheta^2-2\delta\vartheta\rho^2)}{(\rho^2-\vartheta^2)^2}>\frac{\rho^2(\rho^2+\vartheta^2-2\vartheta\rho)}{(\rho^2-\vartheta^2)^2}=\left(\frac{\rho}{\rho+\vartheta}\right)^2>0.$$

Now to prove Proposition 5, it is sufficient to show that ϑ is increasing in γ. Note that

$$C\left(\frac{1}{\rho}\right)=\frac{\tau_u(1-\rho\beta-\rho\gamma)}{\tau_\eta\rho^3}>0 \qquad \text{and} \qquad C\left(\frac{1}{\lambda}\right)=-\frac{\tau_u}{\tau_\eta}\frac{\gamma\beta}{\rho\lambda^2}<0.$$

By the continuity of $C(z)$, it must be the case that $C(z)$ admits a root between $\frac{1}{\rho}$ and $\frac{1}{\lambda}$. Recall from the proof of Proposition 2, ϑ^{-1} is the only outside root, and it follows that $\lambda<\vartheta<\rho$. It also implies that $C(z)$ is decreasing in z in the neighborhood of $z=\vartheta^{-1}$, a property that we use in the sequel to characterize comparative statics of ϑ.

Next, using the definition of $C(z)$, namely

$$C(z)\equiv -z^3+\left(\rho+\frac{1}{\rho}+\frac{1}{\rho}\frac{\tau_u}{\tau_\eta}+\beta\right)z^2-\left(1+\beta\left(\rho+\frac{1}{\rho}\right)+\frac{\beta+\gamma}{\rho}\frac{\tau_u}{\tau_\eta}\right)z+\beta,$$

taking its derivative with respect to γ, and evaluating that derivative at $z=\vartheta^{-1}$, we obtain

$$\frac{\partial C(\vartheta^{-1})}{\partial\gamma}=-\frac{\tau_u}{\rho\tau_\eta}<0.$$

Combining this with the earlier observation that $\frac{\partial C(\vartheta^{-1})}{\partial z}<0$, and using the Implicit Function Theorem, we infer that ϑ is an increasing function of γ.

Similarly, taking derivative with respect to τ_u, we have

$$\frac{\partial C(\vartheta^{-1})}{\partial\tau_u}=\frac{1}{\rho\tau_\eta}\vartheta^{-1}(\vartheta^{-1}-\beta-\gamma)>\frac{1}{\rho\tau_\eta}\vartheta^{-1}(1-\beta-\gamma)>0.$$

Since $\tau_u=\sigma^{-2}$, we conclude that ϑ is also increasing in σ.

Proof of Proposition 6

Given the law of motion of the aggregate outcome $a_t=\frac{\varphi}{1-\delta\rho}\left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\xi_t$, the average forecasts of a_{t+1} and a_{t+2} can be obtained by applying the Wiener-Hopf prediction formula:

$$\overline{\mathbb{E}}_t[a_{t+1}]=\frac{\varphi}{1-\delta\rho}\left(1-\frac{\lambda}{\rho}\right)\frac{1}{1-\vartheta\lambda}\frac{\rho+\vartheta-\rho\vartheta(L+\lambda)}{(1-\vartheta L)(1-\lambda L)}\xi_t,$$
$$\overline{\mathbb{E}}_t[a_{t+2}]=\frac{\varphi}{1-\delta\rho}\left(1-\frac{\lambda}{\rho}\right)\frac{1}{1-\vartheta\lambda}\left(\frac{\rho+\vartheta-\rho\vartheta(L+\lambda)}{(1-\vartheta L)(1-\lambda L)}(\vartheta+\rho)-\frac{\rho\vartheta(1-\rho\vartheta\lambda L)}{(1-\vartheta L)(1-\lambda L)}\right)\xi_t.$$

The average forecast error and the average forecast revision are defined as

$$\text{Error}_t \equiv a_{t+1} - \overline{\mathbb{E}}_t[a_{t+1}], \quad \text{Revision}_t = \overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}],$$

and it follows that

$$\begin{aligned}
\text{Cov}(\text{Error}_t, \text{Revision}_t) &= \left(\frac{\varphi}{1-\delta\rho}\right)^2 \frac{\lambda\vartheta^2(\rho-\lambda)(1-\rho\vartheta)(\rho+\vartheta-\lambda\rho\vartheta)}{\rho^4(1-\lambda^2)(\vartheta-\lambda)^2(1-\lambda\vartheta)^5} \\
&\quad + \left(\frac{\varphi}{1-\delta\rho}\right)^2 \frac{\lambda(\lambda-\rho)(\rho+\vartheta-\lambda\rho\vartheta)(\lambda\rho+\lambda\vartheta-\rho\vartheta-\lambda^2\rho\vartheta)}{\rho^2(1-\lambda^2)(\vartheta-\lambda)(1-\lambda\vartheta)^2} \\
\text{Var}(\text{Revision}_t) &= \left(\frac{\varphi}{1-\delta\rho}\right)^2 \frac{(\lambda-\rho)^2(\rho+\vartheta-\lambda\rho\vartheta)^2}{\rho^2(1-\lambda^2)(1-\lambda\vartheta)^2}.
\end{aligned}$$

The moment K_{CG} can be computed as

$$K_{\text{CG}} = \frac{\text{Cov}(\text{Error}_t, \text{Revision}_t)}{\text{Var}(\text{Revision}_t)} = \lambda \frac{\vartheta+\rho-\rho\vartheta(\lambda+\vartheta)-\rho\lambda\vartheta(1-\lambda\vartheta)}{(\rho-\lambda)(1-\lambda\vartheta)(\rho+\vartheta-\lambda\rho\vartheta)},$$

which is the formula given in the Proposition.

Consider next the partial derivatives of K_{CG} with respect to λ and ϑ:

$$\frac{\partial K_{\text{CG}}}{\partial \lambda} = \frac{\begin{pmatrix} \theta^4\lambda^2\rho\left(\lambda^2\left(\rho^2+1\right)-4\lambda\rho+\rho^2+1\right)-\theta^3\left(4\lambda^3\rho^3+\lambda^2\left(1-6\rho^2\right)+\rho^2\right), \\ +\theta^2\rho\left(\lambda^2\left(6\rho^2-1\right)-4\lambda\rho-\rho^2+1\right)+2\theta\rho^2(1-2\lambda\rho)+\rho^3 \end{pmatrix}}{(1-\theta\lambda)^2(\rho-\lambda)^2(\theta+\rho-\theta\lambda\rho)^2}. \tag{38}$$

$$\frac{\partial K_{\text{CG}}}{\partial \vartheta} = -\frac{\theta\lambda\left(2\rho(1-\theta\lambda)+\theta\right)}{(1-\theta\lambda)^2(\theta+\rho-\theta\lambda\rho)^2} \tag{39}$$

It is possible to verify that $0 < \lambda < \vartheta < \rho < 1$ implies

$$\frac{\partial K_{\text{CG}}}{\partial \lambda} > 0 > \frac{\partial K_{\text{CG}}}{\partial \vartheta}.$$

Because ϑ increases in γ and λ is invariant in γ, we immediately have that K_{CG} is decreasing in γ, as stated in the Proposition.

What remains is to prove that K_{CG} is increasing in σ. This is complicated because σ has opposing effects via λ and ϑ. The rest of the proof deals with this complication. Because the calculations involved are highly cumbersome, we have done them with the help of the analytical tools in Mathematica.

Because λ is a monotone transformation of σ, we can re-express ϑ as function of λ and take the total derivative of K_{CG} with respect to λ instead of its total derivative with respect to σ. That is, we seek to prove $\frac{dK_{\text{CG}}}{d\lambda} > 0$, where

$$\frac{dK_{\text{CG}}}{d\lambda} = \frac{\partial K_{\text{CG}}}{\partial \lambda} + \frac{\partial K_{\text{CG}}}{\partial \vartheta}\frac{\partial \vartheta}{\partial \lambda}, \tag{40}$$

$\frac{\partial K_{\text{CG}}}{\partial \lambda}$ and $\frac{\partial K_{\text{CG}}}{\partial \vartheta}$ are the partial derivatives obtained above, and $\frac{\partial \vartheta}{\partial \lambda}$ is the derivative of ϑ with respect to λ implied by the solution for ϑ. The latter derivative is obtained by re-expressing the cubic in (17) in terms of λ in place of σ and applying the Implicit Function Theorem. In particular, we first re-write the cubic as follows:

$$\rho(1-\beta\theta)(\theta-\lambda)(1-\theta\lambda)-\gamma\theta^2(\rho-\lambda)(1-\lambda\rho)=0. \tag{41}$$

We then apply the Implicit Function Theorem to obtain

$$\frac{\partial\vartheta}{\partial\lambda}=\frac{\rho(\beta\theta-1)\left(\theta^2-2\theta\lambda+1\right)+\gamma\theta^2\left(-2\lambda\rho+\rho^2+1\right)}{\rho\left(\beta\left(3\theta^2\lambda-2\theta\left(\lambda^2+1\right)+\lambda\right)-2\theta\lambda+\lambda^2+1\right)-2\gamma\theta\left(\lambda^2\rho-\lambda\left(\rho^2+1\right)+\rho\right)}. \tag{42}$$

Next, we solve (41) for γ:

$$\gamma=\Gamma(\theta;\lambda,\beta,\rho)\equiv\frac{\rho(1-\beta\theta)(\theta-\lambda)(1-\theta\lambda)}{\theta^2(\rho-\lambda)(1-\lambda\rho)}. \tag{43}$$

This identifies the value of γ that induces as an equilibrium any given value form ϑ in the admissible range $[\lambda,\rho)$. Replacing this value for γ into (42) allows us to re-express the latter as follows:

$$\frac{\partial\vartheta}{\partial\lambda}=\frac{\theta\left(\lambda^2-1\right)(\beta\theta-1)\left(\theta^2\rho-\theta\left(\rho^2+1\right)+\rho\right)}{(\lambda-\rho)(\lambda\rho-1)\left(\beta\theta^3\lambda-\theta\left(\beta\lambda+\lambda^2+1\right)+2\lambda\right)}. \tag{44}$$

Combining the above with (38), (39), and (40), we obtain the following result:

$$\frac{dK_{\text{CG}}}{d\lambda}=\frac{\begin{pmatrix}\beta\lambda^5\rho^4\theta^6+\beta\lambda^3\rho^4\theta^6-5\beta\lambda^4\rho^3\theta^6-\beta\lambda^2\rho^3\theta^6+\beta\lambda^5\rho^2\theta^6+5\beta\lambda^3\rho^2\theta^6-\beta\lambda^4\rho\theta^6-\beta\lambda^2\rho\theta^6\\ -3\beta\lambda^4\rho^4\theta^5+\beta\lambda^2\rho^4\theta^5+3\beta\lambda^3\rho^3\theta^5+\beta\lambda^2\theta^5+3\beta\lambda^4\rho^2\theta^5-2\beta\lambda^2\rho^2\theta^5-3\beta\lambda^3\rho\theta^5-\lambda^6\rho^4\theta^4\\ -\beta\lambda^5\rho^4\theta^4-2\lambda^4\rho^4\theta^4+4\beta\lambda^3\rho^4\theta^4-\lambda^2\rho^4\theta^4-2\beta\lambda\rho^4\theta^4+\beta\lambda^3\theta^4+5\lambda^5\rho^3\theta^4+3\beta\lambda^4\rho^3\theta^4+8\lambda^3\rho^3\theta^4\\ -3\beta\lambda^2\rho^3\theta^4-\lambda\rho^3\theta^4-\lambda^6\rho^2\theta^4-\beta\lambda^5\rho^2\theta^4-8\lambda^4\rho^2\theta^4-2\beta\lambda^3\rho^2\theta^4-4\lambda^2\rho^2\theta^4+2\beta\lambda\rho^2\theta^4+\rho^2\theta^4\\ +\lambda^5\rho\theta^4-\beta\lambda^4\rho\theta^4+3\lambda^3\rho\theta^4+5\lambda^5\rho^4\theta^3+3\beta\lambda^4\rho^4\theta^3+2\lambda^3\rho^4\theta^3-2\beta\lambda^2\rho^4\theta^3+\lambda\rho^4\theta^3-2\lambda^3\theta^3\\ -13\lambda^4\rho^3\theta^3-7\beta\lambda^3\rho^3\theta^3-6\lambda^2\rho^3\theta^3+4\beta\lambda\rho^3\theta^3+\rho^3\theta^3+\lambda^5\rho^2\theta^3+\beta\lambda^4\rho^2\theta^3+7\lambda^3\rho^2\theta^3+\beta\lambda^2\rho^2\theta^3\\ +2\lambda\rho^2\theta^3+2\lambda^4\rho\theta^3+\beta\lambda^3\rho\theta^3+\lambda^2\rho\theta^3-\beta\lambda\rho\theta^3-\rho\theta^3-9\lambda^4\rho^4\theta^2-3\beta\lambda^3\rho^4\theta^2-\lambda^2\rho^4\theta^2+17\lambda^3\rho^3\theta^2\\ +5\beta\lambda^2\rho^3\theta^2+\lambda\rho^3\theta^2-6\lambda^2\rho^2\theta^2-2\beta\lambda\rho^2\theta^2-2\rho^2\theta^2-2\lambda^3\rho\theta^2+2\lambda\rho\theta^2+7\lambda^3\rho^4\theta+\beta\lambda^2\rho^4\theta+\lambda\rho^4\theta\\ -11\lambda^2\rho^3\theta-\beta\lambda\rho^3\theta-\rho^3\theta+4\lambda\rho^2\theta-2\lambda^2\rho^4+2\lambda\rho^3\end{pmatrix}}{(1-\theta\lambda)(1-\lambda\rho)\left(\beta\lambda\theta^3-\lambda^2\theta-\beta\lambda\theta-\theta+2\lambda\right)(\rho-\lambda)^2(\lambda\rho\theta-\theta-\rho)^2} \tag{45}$$

The proof is then completed by verifying that both the numerator and the denominator are positive.

Consider first the denominator and note that this is a decreasing linear function of β. It is therefore positive if and only if $\beta<\frac{\theta\lambda^2+\theta-2\lambda}{\theta^3\lambda-\theta\lambda}$. Because the latter fraction is decreasing in θ, it is bounded from below by the limit of this fraction as $\vartheta\to\rho\to 1$. Because this limit is 1, which is necessarily higher than β, we have that the denominator is necessarily positive.

Consider next the numerator. This, too, is a decreasing linear function of β. And it is positive if and only if

$$\beta < \beta^{\#} \equiv \frac{\begin{pmatrix} \theta^4\rho\left(\lambda^6\left(\rho^3+\rho\right)-\lambda^5\left(5\rho^2+1\right)+2\lambda^4\rho\left(\rho^2+4\right)-\lambda^3\left(8\rho^2+3\right)+\lambda^2\rho\left(\rho^2+4\right)+\lambda\rho^2-\rho\right) \\ -\theta^3\left(\lambda^5\left(5\rho^4+\rho^2\right)+\lambda^4\left(2\rho-13\rho^3\right)+\lambda^3\left(2\rho^4+7\rho^2-2\right)+\lambda^2\left(\rho-6\rho^3\right)+\lambda\rho^2\left(\rho^2+2\right)+\rho\left(\rho^2-1\right)\right) \\ +\theta^2\rho\left(9\lambda^4\rho^3+\lambda^3\left(2-17\rho^2\right)+\lambda^2\rho\left(\rho^2+6\right)-\lambda\left(\rho^2+2\right)+2\rho\right) \\ -\theta\rho^2\left(7\lambda^3\rho^2-11\lambda^2\rho+\lambda\left(\rho^2+4\right)-\rho\right)+2\lambda\rho^3(\lambda\rho-1) \end{pmatrix}}{\theta\lambda\begin{pmatrix} \theta^5\lambda\rho\left(\lambda^3\left(\rho^3+\rho\right)-\lambda^2\left(5\rho^2+1\right)+\lambda\rho\left(\rho^2+5\right)-\rho^2-1\right)-\theta^4\lambda\left(\rho^2-1\right)\left(3\lambda^2\rho^2-3\lambda\rho-\rho^2+1\right) \\ -\theta^3\left(\lambda^4\left(\rho^4+\rho^2\right)+\lambda^3\left(\rho-3\rho^3\right)+\lambda^2\left(-4\rho^4+2\rho^2-1\right)+3\lambda\rho^3+2\rho^2\left(\rho^2-1\right)\right)+ \\ \theta^2\rho\left(\lambda^3\left(3\rho^3+\rho\right)+\lambda^2\left(1-7\rho^2\right)+\lambda\left(\rho-2\rho^3\right)+4\rho^2-1\right)+\theta\rho^2\left(-3\lambda^2\rho^2+5\lambda\rho-2\right)+\rho^3(\lambda\rho-1) \end{pmatrix}}$$

To verify that the above is necessarily true, we return to condition (43).

Recall that this condition gives the value of γ that induces a given θ as an equilibrium. Using this, the primitive $\beta+\gamma<1$ can be re-expressed as $\beta+\Gamma(\theta;\lambda,\beta,\rho)<1$, or equivalently

$$\beta < b^* \equiv \frac{\theta^2\lambda^2\rho+\theta^2(-\lambda)\rho^2+\theta^2\lambda\rho-\theta^2\lambda+\theta^2\rho-\theta\lambda^2\rho-\theta\rho+\lambda\rho}{\theta^3\lambda\rho-\theta^2\lambda\rho^2-\theta^2\lambda+\theta\lambda\rho}. \tag{46}$$

We thus have that $\beta<b^*$ is necessarily satisfied. If we prove that $b^*\le\beta^{\#}$ is also satisfied, we are done.

Let $F(\lambda,\vartheta,\rho)$ denote difference $\beta^{\#}-b^*$ as a function of (λ,ϑ,ρ); this function is obtained simply by using the definitions of these thresholds. We have used Mathematica to verify numerically that F takes non-negative values over the entire $[0,1]^3$ set, which itself necessarily contains the admissible values of (λ,ϑ,ρ). We conclude that both the numerator and the denominator in (45) are positive, which means that K_{CG} is increasing in λ (equivalently, in σ).

Proof of Proposition 7

The proof follows from the main text.

Proof of Proposition 8

See Appendix D.4.

Proof of Proposition 9

Assume that all agents across groups share the same information structure by receiving a private signal about the interest rate r_t

$$x_{i,g,t} = r_t + u_{i,g,t}, \quad u_{i,g,t} \sim \mathcal{N}(0,\sigma^2).$$

We proceed with a guess-and-verify approach. The conjecture is that the law of motion of the aggregate consumption c_t is given by the following AR(2) process for some scalars b and $\vartheta\in(-1,1)$,

$$c_t = \frac{b}{(1-\vartheta L)(1-\rho L)}\eta_t = b\frac{\rho}{\rho-\vartheta}\xi_t - b\frac{\vartheta}{\rho-\vartheta}\zeta_t.$$

where $\xi_t \equiv \frac{1}{1-\rho L}\eta_t$ and $\zeta_t \equiv \frac{1}{1-\vartheta L}\eta_t$. To simplify the notation, denote $\alpha_g \equiv m_g\phi_g$ and $\beta_g \equiv 1-m_g$. Consider the individual best response in group g

$$\begin{aligned} c_{i,g,t} &= -\mathbb{E}_{i,g,t}[r_t] + b\alpha_g\left(\frac{\rho}{\rho-\vartheta}\mathbb{E}_{i,g,t}[\xi_t] - \frac{\vartheta}{\rho-\vartheta}\mathbb{E}_{i,g,t}[\zeta_t]\right) + \beta_g\mathbb{E}_{i,g,t}[c_{i,g,t+1}] \\ &= \frac{1}{1-\beta_g\rho}\left(-1+b\alpha_g\frac{\rho}{\rho-\vartheta}\right)\mathbb{E}_{i,g,t}[\xi_t] - \frac{1}{1-\beta_g\vartheta}b\alpha_g\frac{\vartheta}{\rho-\vartheta}\mathbb{E}_{i,g,t}[\zeta_t]. \end{aligned}$$

Due to the fact that the signal structure is independent of their group identity, the average expectation across the economy is the same as that within the group. The average forecasts of ξ_t and ζ_t are given by

$$\begin{aligned} \overline{\mathbb{E}}_t[\xi_t] &= \left(1-\frac{\lambda}{\rho}\right)\frac{1}{(1-\rho L)(1-\lambda L)}\eta_t, \\ \overline{\mathbb{E}}_t[\zeta_t] &= \left(1-\frac{\lambda}{\rho}\right)\frac{1-\rho\lambda}{1-\vartheta\lambda}\frac{1}{(1-\vartheta L)(1-\lambda L)}\eta_t, \end{aligned}$$

where λ is defined in equation (33). It follows that the average action of group g is

$$c_{g,t} = \frac{1}{1-\lambda L}\left(1-\frac{\lambda}{\rho}\right)\left\{\frac{1}{1-\beta_g\rho}\left(-1+b\alpha_g\frac{\rho}{\rho-\vartheta}\right)\frac{1}{1-\rho L} - \frac{1}{1-\beta_g\vartheta}b\alpha_g\frac{\vartheta}{\rho-\vartheta}\frac{1}{1-\vartheta L}\right\}\eta_t.$$

The aggregate consumption is a weighted average of the actions across different groups

$$\begin{aligned} c_t &= \sum_g \pi_g c_{g,t}, \\ &= \frac{1}{1-\lambda L}\left(1-\frac{\lambda}{\rho}\right)\left\{\sum_g \pi_g\frac{1}{1-\beta_g\rho}\left(-1+b\alpha_g\frac{\rho}{\rho-\vartheta}\right)\frac{1}{1-\rho L} - \sum_g \pi_g\frac{1}{1-\beta_g\vartheta}b\alpha_g\frac{\vartheta}{\rho-\vartheta}\frac{1}{1-\vartheta L}\right\}\eta_t, \\ &\equiv \frac{1}{1-\lambda L}\left(1-\frac{\lambda}{\rho}\right)\frac{\Delta_1-\Delta_2-(\vartheta\Delta_1-\rho\Delta_2)L}{(1-\rho L)(1-\vartheta L)}, \end{aligned}$$

where

$$\begin{aligned} \Delta_1 &= \sum_g \pi_g\frac{1}{1-\beta_g\rho}\left(-1+b\alpha_g\frac{\rho}{\rho-\vartheta}\right), \\ \Delta_2 &= \sum_g \pi_g\frac{1}{1-\beta_g\vartheta}b\alpha_g\frac{\vartheta}{\rho-\vartheta}. \end{aligned}$$

To verify the conjecture, we need to make sure that the actual outcome follows the same AR(2) process as the conjectured one. By matching coefficients, it has to be that

$$\Delta_1 = \frac{\rho-\lambda}{\vartheta-\lambda}\Delta_2, \tag{47}$$

$$b = \left(1-\frac{\lambda}{\rho}\right)(\Delta_1-\Delta_2). \tag{48}$$

Note that without informational frictions, the aggregate outcome is given by

$$c_t = b^*\xi_t, \quad \text{with} \quad b^* = -\frac{\sum_g \pi_g\frac{1}{1-\beta_g\rho}}{1-\sum_g \pi_g\frac{\alpha_g}{1-\beta_g\rho}}.$$

The consumption under perfect information satisfies the standard Euler equation

$$c_t = -\varsigma r_t + \mathbb{E}_t[c_{t+1}],$$

where $-\varsigma \equiv (1-\rho)b^*$.

Going back to the incomplete-information economy, it follows from (47) and(48) that the scale b is given by

$$b = \left(1 - \frac{\vartheta}{\rho}\right) b^*,$$

and ϑ is the inside root of the following equation

$$C(z) = (1 - z\lambda)(z - \lambda)\rho - z(1 - \lambda\rho)(\rho - \lambda)\sum_g \pi_g \frac{\alpha_g}{1 - \beta_g z}.$$

Therefore, the aggregate consumption under incomplete information follows an AR(2) process, which is the same as the baseline case. The particular form of the impact response captured by b also permits the as-if representation, with ω_f and ω_b now being functions of $\{\pi_g, \phi_g, m_g\}$.

For the two-group case, the variable ϑ is the inside root of the following condition by rewriting $C(z)$ as a polynomial equation

$$\widetilde{C}(z) = (1-(1-m_1)z)(1-(1-m_2)z)(1-z\lambda)(z-\lambda)\rho - z(1-\lambda\rho)(\rho-\lambda)Q,$$

where

$$Q = \pi_1 m_1 \phi_1 (1-(1-m_2)z) + \pi_2 m_2 \phi_2 (1-(1-m_1)z).$$

Denote $\phi_1 = \phi$, and by construction, we have $\phi_2 = \frac{1-\pi_1\phi}{\pi_2}$. It follows that

$$\frac{\partial Q}{\partial \phi} = \pi_1 (m_1 - m_2)(1-z).$$

Note that

$$\widetilde{C}(\lambda) = -\lambda(1-\lambda\rho)(\rho-\lambda)(\pi_1 m_1 \phi_1 (1-(1-m_2)\lambda) + \pi_2 m_2 \phi_2 (1-(1-m_1)\lambda)) < 0$$
$$\widetilde{C}(1) = m_1 m_2 \lambda (1-\rho)^2 > 0.$$

Therefore, $\vartheta \in (\lambda, 1)$ and $\widetilde{C}(z)$ is increasing in the neighborhood of ϑ. When $m_1 > m_2$, $\frac{\partial Q}{\partial \phi}|_{z=\vartheta} > 0$. It follows that ϑ is increasing in ϕ.

Proof of Proposition 10

We first show that if $\beta_g \in (0,1)$ and the spectral radius of $(\mathbf{I}-\boldsymbol{\beta})^{-1}\boldsymbol{\gamma}$ is less than 1, then there exists a unique equilibrium. Recall that the individual's best response is

$$a_{i,g,t} = \varphi_g \mathbb{E}_{i,g,t}[\xi_t] + \beta_g \mathbb{E}_{i,g,t}[a_{i,g,t+1}] + \sum_{j=0}^{n} \gamma_{gk} \mathbb{E}_{i,g,t}[a_{j,t+1}] = \varphi_g \mathbb{E}_{i,g,t}\left[\frac{1}{1-\beta_g L^{-1}}\xi_t + \sum_{j=0}^{n} \frac{\gamma_{gk} L^{-1}}{1-\beta_g L^{-1}} a_{j,t}\right]$$

The aggregate outcome for group g is then

$$a_{g,t} = \varphi_g \overline{\mathbb{E}}_{g,t}\left[\frac{1}{1-\beta_g L^{-1}}\xi_t + \sum_{j=0}^{n} \frac{\gamma_{gk} L^{-1}}{1-\beta_g L^{-1}} a_{j,t}\right].$$

By an abuse of notation, we have

$$\boldsymbol{a}_t = \overline{\mathbb{E}}_t\left[(\mathbf{I}-\boldsymbol{\beta}L^{-1})^{-1}\boldsymbol{\varphi}\xi_t + (\mathbf{I}-\boldsymbol{\beta}L^{-1})^{-1}\boldsymbol{\gamma}L^{-1}\boldsymbol{a}_t\right],$$

where $\overline{\mathbb{E}}_t$ denotes $\begin{bmatrix}\overline{\mathbb{E}}_{1,t} & \dots & \overline{\mathbb{E}}_{n,t}\end{bmatrix}'$. Denote $\widetilde{\boldsymbol{\varphi}} \equiv (\mathbf{I}-\boldsymbol{\beta}\rho)^{-1}\boldsymbol{\varphi}$ and $\boldsymbol{\kappa}(L) \equiv (\mathbf{I}-\boldsymbol{\beta}L^{-1})^{-1}\boldsymbol{\gamma}L^{-1}$. The aggregate outcome $\boldsymbol{a}_t$ has the following representation

$$\boldsymbol{a}_t = \widetilde{\boldsymbol{\varphi}}\overline{\mathbb{E}}_t[\xi_t] + \overline{\mathbb{E}}_t\left[\boldsymbol{\kappa}(L)\widetilde{\boldsymbol{\varphi}}\overline{\mathbb{E}}_t[\xi_t]\right] + \overline{\mathbb{E}}_t\left[\boldsymbol{\kappa}(L)\overline{\mathbb{E}}_t\left[\boldsymbol{\kappa}(L)\widetilde{\boldsymbol{\varphi}}\overline{\mathbb{E}}_t[\xi_t]\right]\right] + \dots$$

The aggregate outcome has a unique solution if the power series above is a stationary process or the variance of $a_{g,t}$ is bounded for all g.

Note that: (1) $\mathrm{Var}(\overline{\mathbb{E}}_t[X]) \ge \mathrm{Var}(\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_{t+k}[X]])$ for $k \ge 0$; (2) $\mathrm{Var}(aX+bY) \le (a\sqrt{\mathrm{Var}(X)} + b\sqrt{\mathrm{Var}(Y)})^2$. To show the variance of $a_{g,t}$ is bounded, it is sufficient to show that $\sum_{k=0}^{\infty}\boldsymbol{\kappa}^k(1)$ is bounded. Since $\boldsymbol{\kappa}(1) = (\mathbf{I}-\boldsymbol{\beta})^{-1}\boldsymbol{\gamma}$, if the spectral radius of $(\mathbf{I}-\boldsymbol{\beta})^{-1}\boldsymbol{\gamma}$ is less than 1, $\sum_{k=0}^{\infty}\boldsymbol{\kappa}^k(1)$ is bounded and $\boldsymbol{a}_t$ is stationary.

Now we show that the aggregate outcomes have to be a linear combination of n different AR(2) processes. The signal for agents in group g is

$$x_{i,g,t} = \mathbf{M}(L)\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{i,g,t}\end{bmatrix}, \qquad \text{with} \qquad \mathbf{M}(L) = \begin{bmatrix}\frac{1}{1-\rho L} & \tau_g^{-\frac{1}{2}}\end{bmatrix}.$$

Similar to the proof of Proposition 2, let $B_g(L)$ denote the fundamental representation of the signal process, which is given by

$$B_g(L) = \tau_g^{-\frac{1}{2}}\sqrt{\frac{\rho}{\lambda_g}}\frac{1-\lambda_g L}{1-\rho L},$$

where λ_g is

$$\lambda_g = \frac{1}{2}\left[\rho + \frac{1}{\rho}(1+\tau_g) - \sqrt{\left(\rho + \frac{1}{\rho}(1+\tau_g)\right)^2 - 4}\right].$$

Denote the policy rule of agents in group g as $h_g(L)$, and the law of motion of the aggregate outcome in group g is $a_{g,t} = \frac{h_g(L)}{1-\rho L}\eta_t$. Agents need to forecast the fundamental, their own future action, the aggregate outcomes in each group, which are given by

$$\mathbb{E}_{i,g,t}[\xi_t] = \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \frac{1}{1-\lambda_g L} x_{i,g,t},$$

$$\mathbb{E}_{i,g,t}\left[a_{k,t+1}\right] = \frac{\lambda_g \tau_g}{\rho} \left(\frac{h_k(L)}{(1-\lambda_g L)(L-\lambda_g)} - \frac{h_k(\lambda_g)(1-\rho L)}{(1-\rho\lambda_g)(L-\lambda_g)(1-\lambda_g L)} \right) x_{i,g,t},$$

$$\mathbb{E}_{i,g,t}\left[a_{i,g,t+1} - a_{g,t+1}\right] = \frac{\lambda_g}{\rho} \left(\frac{h_g(L)(L-\rho)}{L(L-\lambda_g)} - \frac{h(\lambda_g)(\lambda_g-\rho)}{\lambda_g(L-\lambda_g)} - \frac{\rho}{\lambda_g} \frac{h_g(0)}{L} \right) \frac{1-\rho L}{1-\lambda_g L} x_{i,g,t}.$$

Using the best response, the fixed point problem is

$$\begin{aligned} h_g(L) x_{i,g,t} = {} & \varphi_g \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \frac{1}{1-\lambda_g L} x_{i,g,t} + \beta_g \frac{\lambda_g}{\rho} \left(\frac{h_g(L)(L-\rho)}{L(L-\lambda_g)} - \frac{h_g(\lambda_g)(\lambda_g-\rho)}{\lambda_g(L-\lambda_g)} - \frac{\rho}{\lambda_g} \frac{h_g(0)}{L} \right) \frac{1-\rho L}{1-\lambda_g L} x_{i,g,t} \\ & + \sum_k \gamma_{g,k} \frac{\lambda_g \tau_g}{\rho} \left(\frac{h_k(L)}{(1-\lambda_g L)(L-\lambda_g)} - \frac{h_k(\lambda_g)(1-\rho L)}{(1-\rho\lambda_g)(L-\lambda_g)(1-\lambda_g L)} \right) x_{i,g,t} \\ & + \beta_g \frac{\lambda_g \tau_g}{\rho} \left(\frac{h_g(L)}{(1-\lambda_g L)(L-\lambda_g)} - \frac{h_g(\lambda_g)(1-\rho L)}{(1-\rho\lambda_g)(L-\lambda_g)(1-\lambda_g L)} \right) x_{i,g,t}. \end{aligned}$$

The system of equation in terms of $\boldsymbol{h}(L)$ is

$$\mathbf{A}(L)\boldsymbol{h}(L) = \boldsymbol{d}(L),$$

where

$$\mathbf{A}(L) = \operatorname{diag}\Big\{L(L-\lambda_g)(1-\lambda_g L)\Big\} - \boldsymbol{\beta} \operatorname{diag}\Big\{\frac{\lambda_g}{\rho}(L-\rho)(1-\rho L) + \frac{\lambda_g \tau_g}{\rho} L\Big\} - \operatorname{diag}\Big\{\frac{\lambda_g \tau_g}{\rho} L\Big\}\boldsymbol{\gamma},$$

and

$$\begin{aligned} d_g(L) = {} & \varphi_g \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} L(L-\lambda_g) - \beta_g (L-\lambda_g)(1-\rho L) h_g(0) \\ & - \left(\beta_g h_g(\lambda_g) \left(\frac{\lambda_g - \rho}{\rho} + \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \right) + \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \sum_k \gamma_{g,k} h_k(\lambda_g) \right) L(1-\rho L). \end{aligned}$$

The solution is given by

$$\boldsymbol{h}(L) = \frac{\operatorname{adj} \mathbf{A}(L)}{\det \mathbf{A}(L)} \boldsymbol{d}(L).$$

Utilizing the identify that

$$\lambda_g + \frac{1}{\lambda_g} = \rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2},$$

the matrix $\mathbf{A}(L)$ can be simplified to

$$\begin{aligned} \mathbf{A}(L) = {} & \operatorname{diag}\Big\{ -\lambda_g L \left(L - \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2} \right) L + 1 \right) \Big\} \\ & + \boldsymbol{\beta} \operatorname{diag}\Big\{ \lambda_g \left(L - \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2} \right) L + 1 \right) \Big\} - \operatorname{diag}\Big\{ \frac{\lambda_g \tau_g}{\rho} L \Big\} \boldsymbol{\gamma}. \end{aligned}$$

The roots of $\det \mathbf{A}(z)$ is the same as the roots of

$$C(z) = \det\left((\boldsymbol{\delta} - \boldsymbol{\gamma} - \mathbf{I}z)\, \mathbf{diag}\left\{ z^2 - \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2}\right) z + 1 \right\} - z\, \mathbf{diag}\left\{ \frac{1}{\rho\sigma_g^2} \right\} \boldsymbol{\gamma} \right).$$

Note that the degree of $\det \mathbf{A}(L)$ is $3n$. Denote the inside roots of $\det \mathbf{A}(L)$ as $\{\zeta_1, \ldots, \zeta_{n_1}\}$ and the outside roots as $\{\theta_1^{-1}, \ldots, \theta_{n_2}^{-1}\}$. Because agents cannot use future signals, the inside roots have to removed. Note that the number of free constants in $\boldsymbol{d}(L)$ is $2n$:

$$\{h_g(0)\}_{g=1}^n, \quad \text{and} \quad \left\{ \beta_g h_g(\lambda_g) \left(\frac{\lambda_g - \rho}{\rho} + \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \right) + \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \sum_k \gamma_{g,k} h_k(\lambda_g) \right\}_{g=1}^n . \tag{49}$$

With a unique solution, it has to be the case that the number of outside roots is n. Also note that by Cramer's rule, $h_g(L)$ is given by

$$h_g(L) = \frac{\det\begin{bmatrix} A_1(L) & \ldots & A_{g-1}(L) & d(L) & A_g(L) & \ldots & A_n(L) \end{bmatrix}}{\det \mathbf{A}(L)}.$$

The degree of the numerator is $3n-1$ as the highest degree of $\boldsymbol{d}_g(L)$ is 1 degree less than that of $A_{g,g}(L)$. By choosing the constants in equation (49), the $2n$ inside roots will be removed. Therefore, the $2n$ constants are solutions to the following system of linear equations:[33]

$$\det\begin{bmatrix} A_1(\zeta_i) & \ldots & A_{g-1}(\zeta_i) & d(\zeta_i) & A_g(\zeta_i) & \ldots & A_n(\zeta_i) \end{bmatrix} = 0, \quad \text{for } i = 1, \ldots, n.$$

After removing the inside roots in the denominator, the degree of the numerator is $n-1$ and the degree of the denominator is n. As a result, the solution to $h_g(L)$ takes the following form

$$h_g(L) = \frac{1}{\Pi_{k=1}^n (1 - \vartheta_k L)} \sum_{k=1}^n \widetilde{\psi}_{g,k} L^{k-1} = \sum_{k=1}^n \psi_{g,k} \left(1 - \frac{\vartheta_k}{\rho}\right) \frac{1}{1 - \vartheta_k L}.$$

In the special case where $\boldsymbol{\beta} = 0$ and $\sigma_g = \sigma$, we have

$$\boldsymbol{a}_t = \boldsymbol{\varphi} \overline{\mathbb{E}}_t[\xi_t] + \boldsymbol{\gamma} \overline{\mathbb{E}}_t[a_{t+1}].$$

Denote the eigenvalue decomposition of $\boldsymbol{\gamma}$ as

$$\boldsymbol{\gamma} \equiv \mathbf{Q}^{-1} \boldsymbol{\Lambda} \mathbf{Q},$$

where $\boldsymbol{\Lambda} = \mathbf{diag}\{\mu_1, \ldots, \mu_n\}$ is a diagonal matrix, and where δ_g is the g-th eigenvalue of $\boldsymbol{\gamma}$. It follows that

$$\mathbf{Q}\boldsymbol{a}_t = \mathbf{Q}\boldsymbol{\varphi} \overline{\mathbb{E}}_t[\xi_t] + \boldsymbol{\Lambda} \overline{\mathbb{E}}_t[\mathbf{Q} a_{t+1}].$$

Denote $\widetilde{\boldsymbol{a}}_t \equiv \mathbf{Q}\boldsymbol{a}_t$. Because $\boldsymbol{\Lambda}$ is a diagonal matrix, it follows that $\widetilde{a}_{g,t}$ is independent of $\widetilde{a}_{j,t}$ for $g \neq j$, and $\widetilde{a}_{g,t}$ satisfies Proposition 2. The degree of complementarity for $\widetilde{a}_{g,t}$ is μ_g, and the corresponding ϑ_g is

[33] The set of constants that solve the system of equations for $h_g(L)$ also solves that for $h_j(L)$ where $i \neq g$. This is because $\{\zeta_i\}_{i=1}^n$ are the roots of the determinant of $\mathbf{A}(L)$, leaving the vectors in $\mathbf{A}(\zeta_i)$ being linearly dependent.

the reciprocal of the outside root of the following quadratic equation:

$$C_g(z) = -z^2 + \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + \beta_g\right) z - \left(1 + \beta_g\left(\rho + \frac{1}{\rho}\right) + \frac{\beta_g + \mu_g}{\rho\sigma^2}\right).$$

Because $\boldsymbol{a}_t$ is a linear transformation of $\widetilde{\boldsymbol{a}}_t$, they share the same AR roots.

Proof of Proposition 11

Now we move to show there exists $\boldsymbol{\omega}_f$ and $\boldsymbol{\omega}_b$ in the complete-information model to rationalize the incomplete-information model solution. In the incomplete-information economy, the average action in group g, $a_{g,t}$, is given by

$$a_{g,t} = \sum_{k=1}^{n} \psi_{g,k}\left(1 - \frac{\vartheta_k}{\rho}\right)\frac{1}{1 - \vartheta_k L}\xi_t.$$

Let $\theta_{k,t} \equiv \left(1 - \frac{\vartheta_k}{\rho}\right)\frac{1}{1-\vartheta_k L}\xi_t$, and it follows that

$$a_{g,t} = \sum_{k=1}^{n} \psi_{g,k}\theta_{k,t}.$$

Denote $\mathbf{Q}$, $\mathbf{\Lambda}$, and $\mathbf{D}$ as

$$\mathbf{Q} \equiv \begin{bmatrix} \psi_{1,1} & \cdots & \psi_{1,n} \\ \vdots & \ddots & \vdots \\ \psi_{n,1} & \cdots & \psi_{n,n} \end{bmatrix}, \qquad \mathbf{\Lambda} \equiv \begin{bmatrix} \vartheta_1 & & \\ & \ddots & \\ & & \vartheta_n \end{bmatrix}, \qquad \mathbf{D} \equiv \begin{bmatrix} 1 - \frac{\vartheta_1}{\rho} \\ \vdots \\ 1 - \frac{\vartheta_n}{\rho} \end{bmatrix}.$$

The vector that collects $\theta_{k,t}$ can be written as

$$\boldsymbol{\theta}_t \equiv \begin{bmatrix} \theta_{1,t} \\ \vdots \\ \theta_{n,t} \end{bmatrix} = \mathbf{\Lambda}\boldsymbol{\theta}_{t-1} + \mathbf{D}\xi_t,$$

and the vector $\boldsymbol{a}_t$ that collects $a_{g,t}$ is

$$\boldsymbol{a}_t = \mathbf{Q}\boldsymbol{\theta}_t = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}\boldsymbol{a}_{t-1} + \mathbf{Q}\mathbf{D}\xi_t.$$

Define $\mathbf{A} \equiv \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ and $\mathbf{B} \equiv \mathbf{Q}\mathbf{D}$, we have

$$\boldsymbol{a}_t = \mathbf{A}\boldsymbol{a}_{t-1} + \mathbf{B}\xi_t. \tag{50}$$

In the perfect-information hybrid model, the law of motion of $\boldsymbol{a}_t$ follows

$$\boldsymbol{a}_t = \boldsymbol{\varphi}\xi_t + \boldsymbol{\omega}_f\boldsymbol{\delta}\mathbb{E}_t[\boldsymbol{a}_{t+1}] + \boldsymbol{\omega}_b\boldsymbol{a}_{t-1}.$$

If (50) is a solution to the perfection-information hybrid model, it has to be that

$$\mathbf{A}\boldsymbol{a}_{t-1}+\mathbf{B}\xi_t=\boldsymbol{\varphi}\xi_t+\boldsymbol{\omega}_f\boldsymbol{\delta}\Big(\rho\mathbf{B}\xi_t+\mathbf{A}(\mathbf{A}\boldsymbol{a}_{t-1}+\mathbf{B}\xi_t)\Big)+\boldsymbol{\omega}_b\mathbf{a}_{t-1}.$$

By method of undetermined coefficients, we have

$$\boldsymbol{\omega}_f\boldsymbol{\delta}(\rho\mathbf{B}+\mathbf{AB})=\mathbf{B}-\boldsymbol{\varphi},$$

$$\boldsymbol{\omega}_b=\mathbf{A}(\mathbf{I}-\boldsymbol{\omega}_f\boldsymbol{\delta}\mathbf{A}).$$

Note that the dimension of $\mathbf{B}-\boldsymbol{\varphi}$ is $n\times 1$ and the dimension of $\boldsymbol{\omega}_f$ is $n\times n$. As a result, ω_f is not uniquely determined.

B The Role of Public Information

Throughout the main analysis, we have assumed that the noise is entirely idiosyncratic. We have thus assumed away, not only correlated errors in expectations, but also the coordination afforded when agents condition their behavior on noisy but public information (Morris and Shin, 2002). In this appendix, we accommodate these possibilities by letting agents observe a public signal in addition to their private signals. We first explain how this modifies our observational equivalence result. We then explain how this matters for our mapping between the theory and the expectations evidence.

B.1 Solution with a Public Signal

In addition to the private signal $x_{i,t}=\xi_t+u_{i,t}$ considered so far, a public signal of the form

$$z_t=\xi_t+\epsilon_t, \tag{51}$$

where $u_{i,t}\sim\mathcal{N}(0,\sigma_u^2)$ and $\epsilon_t\sim\mathcal{N}(0,\sigma_\epsilon^2)$ are, respectively, idiosyncratic and aggregate noises. We next let $\sigma^{-2}\equiv\sigma_u^{-2}+\sigma_\epsilon^{-2}$ measure the overall precision of the available information about the fundamental and $\chi\equiv\frac{\sigma_\epsilon^{-2}}{\sigma_u^{-2}+\sigma_\epsilon^{-2}}$ the fraction of it that reflects public information, or common knowledge.[34]

Proposition 12. *In the extension with public signals described above, the following properties are true.*

(i) The equilibrium outcome is given by

$$a_t=a_t^\xi+v_t,$$

where a_t^ξ is the projection of a_t on the history of ξ_t and v_t is the residual.

[34] It is worth emphasizing that a "public signal" in the theory represents a piece of information that is not only available in the public domain but also common knowledge: every agent observes and acts on it, every agent knows that every other agent observes and acts on it, and so on. Such a signal is therefore at odds with the primary motivation of our paper. It may also not have an obvious empirical counterpart. For instance, aggregate statistics could be effectively observed with idiosyncratic noise due to rational inattention. Nevertheless, the incorporation of a perfect, common-knowledge public signal allows us to shed additional light on the mechanics of the theory as well as on its empirical implications.

(ii) a_t^ξ satisfies Propositions 2 and 3, modulo the replacement of the cubic seen in condition (17) with the following:

$$C(z) \equiv -z^3 + \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + (\delta-\gamma)\right) z^2 - \left(1 + (\delta-\gamma)\left(\rho + \frac{1}{\rho}\right) + \frac{\delta - \gamma\chi}{\rho\sigma^2}\right) z + (\delta-\gamma). \tag{52}$$

(iii) Provided $\gamma > 0$, ϑ is decreasing χ and, therefore, both ω_f and ω_b get closer to their frictionless counterparts as χ increases.

(iv) The residual ν_t follows an AR(1) process with innovation ϵ_t, the noise in the public signal.

Part (i) expresses the equilibrium outcome as the sum of two components: a "fundamental component," defined by the projection of a_t on the history of ξ_t; and a residual, itself measurable in the history of ϵ_t, the aggregate noise.

Part (ii) verifies that all our earlier results extend to the fundamental component here. In other words, although the aggregate outcome is now contaminated by noise, our earlier results continue to characterize its impulse response function (IRF) with respect to the fundamental. Part (ii) also provides the modified cubic that pins down ϑ (and, thereby, the distortions ω_f and ω_b). The old cubic is readily nested in the new one by setting $\chi = 0$.

Part (iii) highlights that, holding σ constant, an increase in χ maps to a smaller ϑ and, thereby, to smaller distortions, but only if $\gamma > 0$; if instead $\gamma = 0$, χ is irrelevant. To understand why, note that an increase in χ for given σ means a substitution of private for public information. This maps to a smaller and less persistent wedge between first- and higher-order beliefs holding constant the dynamics of the first-order beliefs. By the same token, the PE effect of any given innovation remains unchanged, but its GE effect, which is non-zero if and only if $\gamma \neq 0$, is enhanced and gets closer to its frictionless, representative-agent counterpart.

In a nutshell, a higher χ represents an increase in the degree of common knowledge, which in turn amounts to making GE considerations more salient. Clearly, this is a direct extension of the logic developed in our baseline analysis. But what is its empirical content? In particular, does our baseline specification biases upwards the documented distortions by fixing χ at its lowest possible value? As illustrated next, once the theory is required to match relevant evidence on expectations, the incorporation of public information ($\chi > 0$) may actually translate to *higher* distortions than those predicted by our baseline specification ($\chi = 0$).

Part (iv) makes it clear that the residual ν_t is itself an AR(1) transformation of the noise in the public signal. This means that, unlike the fundamental component, the residual does not exhibit hump-shape dynamics.

We find this property is intriguing. If one looks at the response of inflation to either identified monetary shocks (Christiano, Eichenbaum, and Evans, 2005; Romer and Romer, 2004) or to the shock that accounts for most of the business cycle volatility in unemployment, output, or the output gap (the MBC shock in Angeletos, Collard, and Dellas (2019)), one finds a hump shape. But if one looks at the residual, which the DSGE literature captures with a markup shock, then one sees no hump shape. From this perspective, the introduction of public information helps the theory generate a "residual" in inflation that is of the same type as that found in the data. And it helps reconcile why one sees a hump shape in one dimension but not in another.

B.2 Revisiting the Mapping from K_{CG} to (ω_f, ω_b)

Ceteris paribus, the addition of public information reduces the documented distortions by increasing the degree of common knowledge. But it also reduces the predictability of the average forecasts errors. The relevant question is therefore how the accommodation of public information affects the lessons we draw in this paper under the requirement that the theory continues to match the available evidence on expectations.

In our benchmark, which abstracts from public information, the CG coefficient uniquely identifies the value of σ, which in turn pins down the pair (ω_f, ω_b), or equivalently the equilibrium dynamics. Now that we have added a public a signal, the CG coefficient and the equilibrium dynamics alike depend on two unknown parameters, the precisions $\tau_x \equiv \sigma_u^{-2}$ and $\tau_z \equiv \sigma_\epsilon^{-2}$ of, respectively, the private and the public information. As a result, we loose *point* identification but preserve *set* identification: only certain pairs of τ_z and τ_x are consistent, under the lens of the theory, with the evidence in CG. Furthermore, because the theoretical value of K_{CG} converges to zero as the public information becomes sufficiently precise, the estimated value of K_{CG} puts an upper bound on τ_z.[35]

Figure 6 illustrates the implications of these properties for the documented distortions within the context of our application to inflation (Section 6). On the horizontal axis, we let τ_z vary between zero (our benchmark) and the aforementioned bound. For each τ_z in this range, we find the value of τ_x that matches the point estimate of K_{CG} provided in CG and report the implied values for ω_f and ω_b.

For the application under consideration, the upper bound on τ_z turns out to be quite low. This is because evidence in CG points towards considerable predictability in average forecast errors, which in turn requires a significant departure from common knowledge. What is more, the distortions *increase* as we raise τ_z within the admissible range. That is, once the theory is disciplined with the relevant evidence, the incorporation of public information *reinforces* the documented distortions.

Similar points apply if we let for an *endogenous* public signal of the form $z_t = a_t + \epsilon_t$, which in the application under consideration can be thought of as statistic of inflation contaminated with measurement

[35] That is, the set of the admissible values for the pair (τ_x, τ_z) can be expressed as

$$S(K_{\text{CG}}) = \left\{(\tau_x, \tau_z) : \tau_z \leq T(K_{\text{CG}}) \text{ and } \tau_x = f(\tau_z, K_{\text{CG}})\right\},$$

where K_{CG} is the CG moment, $T(\cdot)$ is a function that gives corresponding upper bound on τ_z, and $f(\cdot)$ is a function that gives the value of τ_x that lets the theory match this moment for any given τ_z below the aforementioned bound.

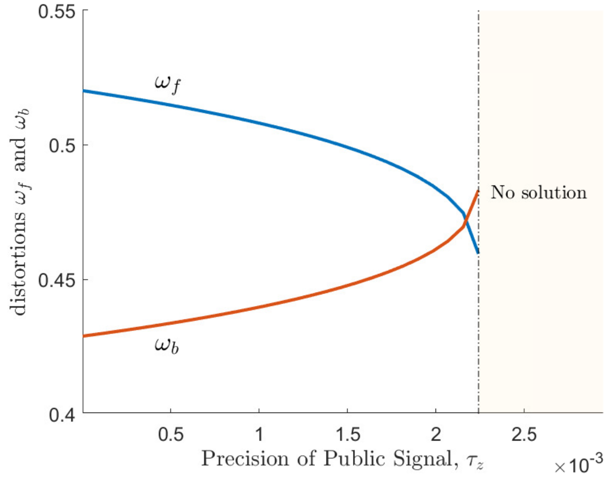

Figure 6: The Role of Public Information

error.[36] Similar to the exogenous-information case, matching the CG moment puts an upper bound on the informativeness of this signal. Different from the exogenous-information case, this informativeness is now endogenous to the actual inflation dynamics. This introduces an additional fixed point problem, which can only be solved numerically. But as illustrated in Figure 7 in Appendix D.3, the main message goes through.

C Idiosyncratic Shocks and Micro- vs Macro-level Distortions

The various adjustment costs assumed in the DSGE literature are supposed to be equally present at the macroeconomic and the microeconomic level. But this is not true. For instance, the macroeconomic estimates of the habit in consumption obtained in the DSGE literature are much larger than the corresponding microeconomic estimates (see Havranek, Rusnak, and Sokolova, 2017, for a metanalysis).

Consider next the menu-cost literature that aims at accounting for the microeconomic data on prices (Golosov and Lucas Jr, 2007; Midrigan, 2011; Alvarez and Lippi, 2014; Nakamura and Steinsson, 2013). Different "details" such as the number of products that are simultaneously re-priced and the so-called selection effect matter for how steep the effective Philips curve is, but do *not* help generate the requisite sluggishness in inflation that the DSGE literature captures with the ad hoc Hybrid NKPC.

A similar point applies to the literature that aims at accounting for the lumpiness of investment at the plant level (Caballero and Engel, 1999; Bachmann, Caballero, and Engel, 2013): this literature has *not* provided support for the kind of adjustment costs to investment employed in the DSGE literature.

In sort, whether one goes "downstream" from DSGE models to their microeconomic implications or

[36] This specification is close to that studied in Nimark (2008). The main difference is that the theory is herein disciplined by the evidence in Coibion and Gorodnichenko (2015).

"upstream" from the more realistic, fixed-cost models used to account for the microeconomic data to their macroeconomic implications, there is a pervasive gap between micro and macro.

Our result that the distortions increase with the importance of GE considerations contributes towards filling this micro-to-macro gap. When an individual responds to aggregate shocks, she has to predict the responses of others and align hers with theirs. To the extent that GE considerations are strong enough, this generates a feedback loop from sluggish expectations to sluggish outcomes and *back*. When instead an individual responds to idiosyncratic shocks, this mechanism is muted. Furthermore, agents may naturally have much more information about idiosyncratic shocks than about aggregate shocks both because of decentralized market interactions (Lucas, 1972) and because of rational inattention Maćkowiak and Wiederholt (2009). It follows that the documented distortions may loom large at the macroeconomic time series even if they appear to be small in the microeconomic time series.

We illustrate this point in the rest of this appendix by adding idiosyncratic shocks to our framework. The optimal behavior of agent i now obeys the following equation:

$$a_{i,t} = \mathbb{E}_{i,t}[\varphi \xi_{i,t} + \beta a_{i,t+1} + \gamma a_{t+1}], \tag{53}$$

where

$$\xi_{i,t} = \xi_t + \zeta_{i,t}.$$

and where $\zeta_{i,t}$ is a purely idiosyncratic shock. We let the latter follow a similar AR(1) process as the aggregate shock: $\zeta_{i,t} = \rho\zeta_{i,t-1} + \epsilon_{i,t}$, where $\epsilon_{i,t}$ is i.i.d. across both i and t.[37]

We then specify the information structure as follows. First, we let each agent observe the same signal $x_{i,t}$ about the aggregate shock ξ_t as in our baseline model. Second, we let each agent observe the following signal about the idiosyncratic shock $\zeta_{i,t}$:

$$z_{i,t} = \zeta_{i,t} + v_{i,t},$$

where $v_{i,t}$ is independent of $\zeta_{i,t}$, of ξ_t, and of $x_{i,t}$.

Because the signals are independent, the updating of the beliefs about the idiosyncratic and the aggregate shocks are also independent. Let $1 - \frac{\lambda}{\rho}$ be the Kalman gain in the forecasts of the aggregate fundamental, that is,

$$\mathbb{E}_{i,t}[\xi_t] = \lambda \mathbb{E}_{i,t-1}[\xi_t] + \left(1 - \frac{\lambda}{\rho}\right) x_{i,t}.$$

Next, let $1 - \frac{\hat{\lambda}}{\rho}$ be the Kalman gain in the forecasts of the idiosyncratic fundamental, that is,

$$\mathbb{E}_{i,t}[\zeta_{i,t}] = \hat{\lambda} \mathbb{E}_{i,t-1}[\zeta_{i,t}] + \left(1 - \frac{\hat{\lambda}}{\rho}\right) z_{i,t}.$$

[37] The restriction that the two kinds of shocks have the same persistence is only for expositional simplicity.

It is straightforward to extend the results of Section 3.3 to the current specification. It can thus be shown that the equilibrium action is given by the following:

$$a_{i,t} = \left(1 - \frac{\hat{\lambda}}{\rho}\right) \frac{\varphi}{1-\rho\beta} \frac{1}{1-\hat{\lambda}L} \zeta_{i,t} + \left(1 - \frac{\vartheta}{\rho}\right) \frac{\varphi}{1-\rho\delta} \frac{1}{1-\vartheta L} \xi_t + u_{i,t},$$

where ϑ is determined in the same manner as in our baseline model and where $u_{i,t}$ is a residual that is orthogonal to both $\zeta_{i,t}$ and ξ_t and that captures the combined effect of all the idiosyncratic noises in the information of agent i. Finally, it is straightforward to check that $\vartheta = \lambda$ when $\gamma = 0$; $\vartheta > \lambda$ when $\gamma > 0$; and the gap between ϑ and λ increases with the strength of the GE effect, as measured with γ.

In comparison, the full-information equilibrium action is given by

$$a^*_{i,t} = \frac{\varphi}{1-\rho\beta} \zeta_{i,t} + \frac{\varphi}{1-\rho\delta} \xi_t.$$

It follows that, relative to the full-information benchmark, the distortions of the micro- and the macro-level IRFs are given by, respectively,

$$\left(1 - \frac{\hat{\lambda}}{\rho}\right) \frac{1}{1-\hat{\lambda}L} \qquad \text{and} \qquad \left(1 - \frac{\vartheta}{\rho}\right) \frac{1}{1-\vartheta L}.$$

The macro-level distortions is therefore higher than its micro-level counterpart if and only if $\vartheta > \hat{\lambda}$.

As already mentioned, it is natural to assume that $\hat{\lambda}$ is lower than λ, because the typical agent is likely to be better informed about, allocate more attention to, idiosyncratic shocks relative to aggregate shocks. This guarantees a lower distortion at the micro level than at the macro level even if we abstract from GE interactions (equivalently, from higher-order uncertainty). But once such interactions are taken into account, we have that ϑ remains higher than $\hat{\lambda}$ even if $\hat{\lambda} = \lambda$. That is, even if the first-order uncertainty about the two kind of shocks is the same, the distortion at the macro level may remain larger insofar as there are positive GE feedback effects, such as the Keynesian income-spending multiplier or the dynamic strategic complementarity in price-setting decisions of the firms.

In short, the mechanism identified in our paper is distinct from the one identified in Maćkowiak and Wiederholt (2009) and employed in subsequent works such as Carroll et al. (2020) and Zorn (2018), but the two mechanisms complement each other towards generating more pronounced distortions at the macro level than at the micro level. The two mechanisms are combined in recent work by Auclert, Rognlie, and Straub (2020).

D Application to Inflation: Micro-foundations and Additional Results

D.1 Derivation of Incomplete-Information NKPC

The original derivations of the incomplete-information versions of the Dynamic IS and New Keynesian Philips curves seen in conditions (8) and (9) can be found in Angeletos and Lian (2018). Those derivations are based in an extension of the New Keynesian model that incorporates a variety of idiosyncratic and aggregate shocks so as to noise up the information that consumers and firms may extract from the perfect observation of concurrent prices, wages, and other endogenous outcomes. Here, we offer a simplified derivation that bypasses these "details" and, instead, focuses on the essence. To economize, we do so only in the context of the NKPC, which is the application we push quantitatively. We also use this as an opportunity to point out a mistake in the variant equations found in Nimark (2008) and Melosi (2016).

Apart for the introduction of incomplete information, the micro-foundations are the same as in familiar textbook treatments of the NKPC (e.g., Galí, 2008). There is a continuum of firms, each producing a differentiated commodity. Firms set prices optimally, but can adjust them only infrequently. Each period, a firm has the option to reset its price with probability $1-\theta$, where $\theta \in (0,1)$; otherwise, it is stuck at the previous-period price. Technology is linear, so that the real marginal cost of a firm is invariant to its production level.

The optimal reset price solves the following problem:

$$P^*_{i,t} = \arg\max_{P_{i,t}} \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,t}\left\{Q_{t|t+k}\left(P_{i,t}Y_{i,t+k|t} - P_{t+k}\mathtt{mc}_{t+k}Y_{i,t+k|t}\right)\right\}$$

subject to the demand equation, $Y_{i,t+k} = \left(\frac{P_{i,t}}{P_{t+k}}\right)^{-\epsilon} Y_{t+k}$, where $Q_{t|t+k}$ is the stochastic discount factor between t and $t+k$, Y_{t+k} and P_{t+k} are, respectively, aggregate income and the aggregate price level in period $t+k$, $P_{i,t}$ is the firm's price, as set in period t, $Y_{i,t+k|t}$ is the firm's quantity in period $t+k$, conditional on not having changed the price since t, and $\mathtt{mc}_{t+k}$ is the real marginal cost in period $t+k$.

Taking the first-order condition and log-linearizing around a steady state with no shocks and zero inflation, we get the following, familiar, characterization of the optimal rest price:

$$p^*_{i,t} = (1-\chi\theta)\sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,t}[\mathtt{mc}_{t+k} + p_{t+k}]. \tag{54}$$

We next make the simplifying assumption that the firms observe that past price level but do not extract information from it. Following Vives and Yang (2017), this assumption can be interpreted as a form of bounded rationality or inattention. It can also be motivated on empirical grounds: in the data, inflation contains little statistical information about real marginal costs and output gaps—it's dominated by the residual, or what the DSGE literature interprets as "markup shocks." This means that, even if we were to allow firms to extract information from past inflation, this would make little quantitative difference, provided that we accommodate an empirically relevant source of noise. Furthermore, as we show in the end of Section 6, our observational-equivalence result remains a useful approximation of the true equilibrium in extension that allow for such endogenous information.

With this simplifying assumption, we can restate condition (54) as

$$p_{i,t}^{*} - p_{t-1} = (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^{k}\mathbb{E}_{i,t}[\mathtt{mc}_{t+k}] + \sum_{k=0}^{\infty}(\chi\theta)^{k}\mathbb{E}_{i,t}[\pi_{t+k}], \tag{55}$$

Since only a fraction $1-\theta$ of the firms adjust their prices each period, the price level in period t is given by $p_t = (1-\theta)\int p_{i,t}^{*}di + \theta p_{t-1}$. By the same token, inflation is given by

$$\pi_t \equiv p_t - p_{t-1} = (1-\theta)\int\left(p_{i,t}^{*} - p_{t-1}\right).$$

Combining this with condition (55) and rearranging, we arrive at the following expression:

$$\pi_t = \kappa\sum_{k=0}^{\infty}(\chi\theta)^{k}\overline{\mathbb{E}}_t[\mathtt{mc}_{t+k}] + \chi(1-\theta)\sum_{k=0}^{\infty}(\chi\theta)^{k}\overline{\mathbb{E}}_t[\pi_{t+k+1}]. \tag{56}$$

where $\kappa \equiv \frac{(1-\chi\theta)(1-\theta)}{\theta}$. This is the same as condition 25 in the main text.

When information is complete, we can replace $\overline{\mathbb{E}}_t[\cdot]$ with $\mathbb{E}_t[\cdot]$, the expectation of the representative agent. We can then use the Law of Iterated Expectations to reduce condition (56) to the standard NKPC. When instead information is incomplete, the Law of Iterated Expectations does not apply at the aggregate level, because average forecast errors can be auto-correlated, and therefore condition (56) cannot be reduced to the standard NKPC.

As explained in the main text, condition (56) involves extremely complex higher-order beliefs and precludes a sharp connection to the data—and this is where the toolbox provided in this paper comes to rescue.

Let us now explain the two reasons why the incomplete-information NKPC seen in condition (56) is different from that found in Nimark (2008) and Melosi (2016). The first reason is that, while we let firms observe the current-period price level, these papers let them observe only the past-period price level. Clearly, this difference vanishes as the time length of a period gets smaller. The second, and most important, reason is a mistake, which we explain next.

Take condition (54) and rewrite it in recursive form as follows:

$$p_{i,t}^{*} = (1-\chi\theta)\mathbb{E}_{i,t}[\mathtt{mc}_t + p_t] + (\chi\theta)\mathbb{E}_{i,t}[p_{i,t+1}^{*}].$$

Aggregate this condition yields a term of the form $\int\mathbb{E}_{i,t}[p_{i,t+1}^{*}]di$, the average expectation of the *own* reset price, in the right-hand side. And this is where the oversight occurs: the aforementioned term is inadvertently replaced with the average expectation of the *average* reset price.

In more abstract terms, this is like equating $\int\mathbb{E}_{i,t}[a_{i,t+1}]di$ with $\int\mathbb{E}_{i,t}[a_{t+1}]di$. If this were true, we could have readily aggregated condition (4) to obtain the following equation:

$$a_t = \varphi\overline{\mathbb{E}}_t[\xi_t] + \delta\overline{\mathbb{E}}_{t+1}[a_{t+1}].$$

Relative to condition (5), this amounts to dropping the expectations of the aggregate outcome a horizons $k \geq 2$, or restricting $\beta = 0$. But this is not true. Except for knife-edge cases such as that of an improper prior, incomplete information implies that the typical agent forms a different expectation about his own actions than the actions of others, which means that

$$\int \mathbb{E}_{i,t}[a_{i,t+1}]di \neq \int \mathbb{E}_{i,t}[a_{t+1}]di.$$

and the aforementioned simplification does not apply.

D.2 Decomposition of PE and GE in Figure 3

This appendix describes the construction of the dotted red line in Figure 3, that is, the counterfactual that isolates the PE channel. This builds on the decomposition between PE and GE effects first introduced in in Section 3.1.

Using condition (56), the incomplete-information inflation dynamics can be decomposed into two components: the belief of the present discounted value of real marginal costs, $\varphi\sum_{k=0}^{\infty}\beta^k\overline{\mathbb{E}}_t[\mathtt{mc}_{t+k}]$; and the belief of of the present discounted value of inflation, $\gamma\sum_{k=0}^{\infty}\beta^k\overline{\mathbb{E}}_t[\pi_{t+k+1}]$. The same decomposition can also be applied when agents have perfect information:

$$\pi_t^* = \underbrace{\varphi\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t[\mathtt{mc}_{t+k}|\mathtt{mc}_t]}_{\text{complete-info PE component}} + \underbrace{\gamma\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t\left[\pi^*_{t+k+1}|\mathtt{mc}_t\right]}_{\text{complete-info GE component}}. \tag{57}$$

A natural question is which component contributes more to the anchoring of inflation as we move from the complete to incomplete information.

To answer this question, we define the following auxiliary variable:

$$\widetilde{\pi}_t = \underbrace{\varphi\sum_{k=0}^{\infty}\beta^k\overline{\mathbb{E}}_t[\mathtt{mc}_{t+k}]}_{\text{incomplete-info PE component}} + \underbrace{\gamma\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t\left[\pi^*_{t+k+1}|\mathtt{mc}_t\right]}_{\text{complete-info GE component}}. \tag{58}$$

The difference between π_t^* and $\widetilde{\pi}_t$ measures the importance of beliefs about real marginal costs, and the difference between $\widetilde{\pi}_t$ and π_t measures the importance of beliefs about inflation.

The dotted red line in Figure 3 corresponds to $\widetilde{\pi}_t$. Clearly, most of the difference between complete and incomplete information is due the anchoring of beliefs about future inflation. Or, to put it in terms of our discussion of PE and GE effects, most of the action is through the GE channel.

The logic behind this finding can be understood by computing the GE multiplier that is hidden inside the standard NKPC. Let μ^* be the ratio of the GE component to the PE component under complete information, that is, the ratio of the two terms seen in condition (57). This identifies the GE multiplier; the total effect is $1+\mu^*$ times the PE effect. Straightforward calculation shows that

$$\mu^* = \frac{\rho\chi(1-\theta)}{1-\chi\rho} \approx 6.4.$$

That is, even in the familiar, complete-information benchmark, the expectations of future inflation are 6.4 times more important than the expectations of future real marginal costs in driving actual inflation. This in turn helps explains why most of the informational friction works through the GE channel, or the anchoring of the expectations of inflation, as seen in Figure 3 in the main text.

D.3 Adding Public Information

In Section 6, we quantified the effects of the informational friction assuming away public information. Here, building on the insights developed in Appendix B, we illustrate how that exercise has provided a *conservative* estimate of the effects that are obtained once we add public information. We further show that this point is reinforced if the public information is endogenous.

We thus consider two cases: an *exogenous* public signal of the form $z_t = \text{mc}_t + \text{noise}$, and an *endogenous* public signal of the form $z_t = \pi_t + \text{noise}$, namely a noisy statistic of inflation. The first case affords an analytical characterization, along the lines of Appendix B; the second case requires a numerical approximation but, as shown below, only reinforces our message.[38]

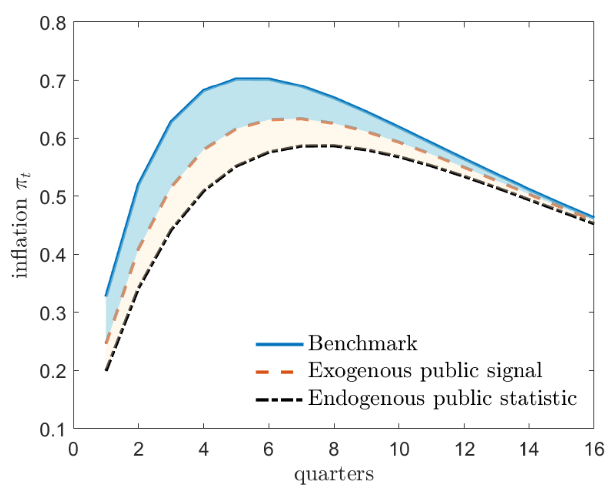

Figure 7: IRF of Inflation, Exogenous vs Endogenous Information

Figure 7 compares the IRF of inflation to innovations in the real marginal cost under three information structures, all required to match the regression coefficient K_{CG} estimated in CG. The blue, solid line

[38]We thank an anonymous referee for suggesting these explorations.

corresponds to our benchmark, which abstracts from public information. As explained in Appendix B, once we allow for a public signal, there is a range of admissible values for its precision, each one mapping to a different pair (ω_f, ω_b), or a different IRF. The red, dashed line in the figure gives the IRF that is obtained when the public signal is exogenous and its precision is the maximal one consistent with K_{CG}. The area between this line and the benchmark line spans all the admissible parameterizations of the exogenous-information case. Finally, the black, dotted line gives the IRF that obtains when the public signal is endogenous and its precision equals the appropriate upper bound. The area between this line and the benchmark line spans all the admissible parameterizations of the endogenous-information case.

The main takeaways are twofold. First, the exogenous-information setting provides a useful analytical tool to understand the more realistic but less tractable endogenous-information case. Second, the accommodation of public information, exogenous or endogenous, only reinforces the quantitative findings once the theory is disciplined by the available evidence on expectations.[39]

D.4 Market Concentration

In the environment where each market consists only a finite number of firms, the (log-linearized) individual firm's optimal reset price is characterized as below.

Lemma 1. *The optimal reset price of individual firm i in market m follows*

$$p^*_{i,m,t} = (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \kappa \mathbb{E}_{i,m,t}[mc_{t+k}] + (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \mathbb{E}_{i,m,t}[\alpha_N p_{m-i,t+k} + (1-\alpha_N)p_{t+k}], \tag{59}$$

where α_N is given by

$$\alpha_N = \frac{N(\psi-1)(\psi-\varepsilon)}{\psi\left(N^2(\psi-1)-(N-1)\psi\right)+(N-2)\psi\varepsilon+\varepsilon^2}.$$

In condition (59), χ, θ, and κ are the same parameters as in the baseline NKPC setup, while $\alpha_N \in (0,1)$ is a new scalar which summarizes how much a firm's pricing strategy depends on the prices of its competitors relative to the aggregate price level. It is easy to verify that $\psi > 1$ and $\psi > \epsilon$ suffices for α_N to be decreasing in N. And in the special case in which $\psi = \infty$, which amounts to a Cournot-like game for each market, we have more simply that $\alpha_N = 1/(2N)$.

The economy-wide inflation can be obtained by aggregating the above condition across markets, which leads to a modified version of our incomplete-information NKPC.

[39] A third, subtler takeaway is that the endogenous public signal contributes to more persistence than the exogenous one. We find this intriguing and we suspect it is because inflation moves more sluggishly than the fundamental, thus slowing down the learning. Nimark (2008) also hypothesizes that endogenous signals add persistence. The logic is, however, complicated by the fact that, as we vary the form of the signal, we adjust its precision to make sure that theory keeps matching the CG moment.

Lemma 2. *The aggregate inflation rate follows*

$$\pi_t = \kappa \sum_{k=0}^{\infty} \left(\frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \overline{\mathbb{E}}_t[mc_{t+k}] + \frac{\chi(1-\theta)(1-\alpha_N)}{1-(1-\theta)\alpha_N} \sum_{k=0}^{\infty} \left(\frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \overline{\mathbb{E}}_t[\pi_{t+k+1}]. \tag{60}$$

For our purposes, the key observation is that α_N is decreasing in N, or decreasing in market concentration. Intuitively, as $N \to \infty$ and the firm becomes infinitesimally small not only vis-a-vis the entire economy but also vis-a-vis its own market, the firm only care to set a price in proportion to its nominal marginal cost, which itself is driven by the aggregate price level. That is, as $N \to \infty$, α_N approaches 1, condition (60) reduces to condition (25), and we recover the case studied before. But when N is finite, a new consideration emerges: when a firm raises its price, it depresses its market share. This effect scales up with market concentration, explaining why higher market concentration maps to a higher α_N, or a higher consideration for local conditions relative to aggregate conditions.

Under complete information, this consideration is of no consequence for the aggregate inflation dynamics: when an aggregate shock to the real marginal cost occurs, a typical firm expects both its immediate competitors and the rest of the economy to respond in tandem, so it makes no difference how much firms care about the former versus the latter. But when information is incomplete, and under the plausible assumption that firms know more about their immediate competitors than about the rest of the economy, the aforementioned consideration amounts to reducing the extent of higher-order uncertainty and its footprint on the inflation dynamics.

These points are evident from condition (60). Mapping this condition to our framework yields

$$\gamma = \frac{\chi(1-\theta)(1-\alpha_N)}{1-(1-\theta)\alpha_N} \quad \text{and} \quad \beta = \frac{\chi\theta}{1-(1-\theta)\alpha_N} = \chi - \gamma.$$

That the sum $\beta + \gamma$ equals χ means that, with complete information, inflation continues to obey the standard NKPC ($\pi_t = \kappa \text{mc}_t + \chi \mathbb{E}_t \pi_{t+1}$) and is invariant to market concentration. That γ increases with α_N means that higher market concentration maps to a smaller degree of strategic complementarity and thereby to a smaller ϑ in the incomplete-information outcome. Applying our observational-equivalence result then yields Proposition 8.

E Heterogeneity à la HANK

In this appendix we detail the micro-foundations of the HANK application considered in Section 7. As described in the main text, households are heterogeneous in terms of mortality risk, associated MPC, and exposure to business cycles. They can trade annuities, so as to insure against mortality risk, but are precluded from trading more sophisticated assets such as GDP futures, so that we can bypass the complications of endogenous information aggregation. We also let firms' profits be taxed by the government, and distributed to consumers in proportion to labor income and regardless of age. This makes sure that consumers of all types and ages hold zero financial wealth in steady state. And we shut down the distribution effects of interest-rate shocks by appropriate fiscal transfers, as explained shortly.

Consider a consumer i, of type g, born in period τ. Taking into account the mortality risk, her expected lifetime utility at birth is given by

$$\sum_{t=\tau}^{\infty} \left(\chi \varpi_g\right)^{t-\tau} \log\left(C_{i,g,t;\tau}\right),$$

where $C_{i,g,\tau;t}$ denotes her consumption in period t (conditional on survival) and $\chi \in (0,1)$ is the subjective discount factor. Her budget constraint, on the other hand, is given by

$$C_{i,g,\tau;t} + S_{i,g,\tau;t} = \frac{R_{t-1}}{\varpi_g} S_{i,g,\tau;t-1} + (Y_t)^{\phi_g} + T_{g,t}, \quad \forall \tau \geq t$$

where $S_{i,g,t;\tau}$ denotes savings in terms of the annuity, Y_t denotes aggregate income, $T_{g,t}$ denotes a group-specific lump-sum transfer, and ϕ_g parameterizes the elasticity of group g's income with respect to aggregate income.

We henceforth work with the log-linearized solution around a steady state in which there are no shocks, $\chi R_t = 1$, and $C_t = Y_t = Y^*$, where Y^* is the natural rate of output.[40] We use lower-case variables to represent log-deviations from the steady state (e.g., $r_t \equiv \log R_t - \log \chi^{-1}$), with the exception that $s_{i,g,\tau;t}$ and $\tau_{g,t}$ stand for, respectively, $\frac{S_{i,g,\tau;t}}{Y^*}$ and $\frac{T_{g,t}}{Y^*}$ as their steady-state values are zero. We can then express the optimal expenditure of a consumer in group g as follows:

$$\begin{aligned} c_{i,g,\tau,t} =& (1-\chi\varpi_g)\left(\frac{1}{\chi\varpi_g} s_{i,g,\tau,t-1} + \mathbb{E}_{i,t}[\mathcal{T}_{g,t}]\right) - \chi\omega_g \sum_{j=0}^{\infty} (\chi\varpi_g)^j \mathbb{E}_{i,t}[r_{t+j}] \\ &+ (1-\chi\varpi_g)\phi_g \sum_{j=0}^{\infty} (\chi\varpi_g)^j \mathbb{E}_{i,t}[y_{t+j}] \end{aligned} \tag{61}$$

where $\mathcal{T}_{g,t} \equiv \sum_{j=0}^{\infty} (\chi\varpi_g)^j \tau_{g,t+j}$ captures the present discounted value of transfers.

The average consumption of group g in period t is given by

$$c_{g,t} \equiv (1-\omega_g) \sum_{j=0}^{\infty} (\varpi_g)^j \int c_{i,g,t-j,t} di.$$

Aggregating (61) across all consumers of any given group g, we get

$$\begin{aligned} c_{g,t} = & (1-\chi\omega_g)\left(\frac{1}{\chi} s_{g,t-1} + \overline{\mathbb{E}}_t[\mathcal{T}_{g,t}]\right) - \chi\omega_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[r_{t+j}] \\ &+ (1-\chi\omega_g)\phi_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[y_{t+j}]. \end{aligned} \tag{62}$$

Similarly, by aggregating the budget constraints of all consumers in group g, and taking into account

[40] To simplify the exposition, we suppress the production side of the economy and the determination of the flexible-price outcomes. The details can be filled in the usual way; let technology be linear in labor and assume constant aggregate productivity to get a time-invariant natural rate of output.

how the annuities effectively redistribute wealth from deceased to surviving agents, we get the following group-level budget constraint:

$$c_{g,t} + s_{g,t} = \frac{1}{\chi} s_{g,t-1} + \phi_g y_t - \tau_{g,t},$$

where $s_{g,t}$ is the saving of group g.

Market clearing imposes $y_t = c_t$, or equivalently $s_t = 0$, where $c_t \equiv \sum_g \pi_g c_{g,t}$ and $s_t \equiv \sum_g \pi_g s_{g,t}$. We close the model by specifying a rule for fiscal policy (more on this below) and by treating the real interest rate as an exogenous AR(1) process, with persistence ρ. As mentioned in the main text, this amounts to studying the aggregate-demand effects of a monetary policy that targets such a process for the real interest rate. Alternatively, one can assume that prices are infinitely rigid, in which case r_t coincides with the nominal rate (the policy instrument) and its innovations can be interpreted monetary shocks.

Let us now fill in the details of fiscal policy. For the analysis in the main text, we let the transfers be such that following condition is satisfied in every period:

$$\sum_g \pi_g (1 - \chi\omega_g) s_{g,t} + \sum_g \pi_g \overline{\mathbb{E}}_t[\mathcal{T}_{g,t}] = 0, \tag{63}$$

When all groups have the same MPC (i.e., $\omega_g = \omega_{g'}$ for all g, g'), this condition is trivially satisfied with $\mathcal{T}_{g,t} = 0$ for all g, t. When instead different groups have different MPCs, this condition requires that fiscal policy offsets the interaction of MPC heterogeneity with wealth inequality. In particular, a sufficient condition for (63) to hold is that $\overline{\mathbb{E}}_t[\mathcal{T}_{g,t}] = (1 - \chi\omega_g)s_{g,t}$ for all g, t. And since $s_{g,t}$ is measurable in the history of the aggregate shock alone, the transfers do *not* have to be conditioned on the consumers' age or idiosyncratic histories.

As long as condition (63) is satisfied, we can aggregate condition (62) across groups to obtain the economy-wide aggregate consumption as follows:

$$c_t = \sum_g \pi_g \Big\{ -\chi\omega_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[r_{t+j}] + (1 - \chi\omega_g)\phi_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[y_{t+j}] \Big\} \tag{64}$$

Combining this with market clearing, or $c_t = y_t$, we infer that the equilibrium process of aggregate income (and aggregate consumption) in this economy is the same as the solution of a network where the best response of group g is given by

$$y_{g,t} = -\chi\omega_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[r_{t+j}] + (1 - \chi\omega_g)\phi_g \sum_{j=0}^{\infty} (\chi\omega_g)^j \overline{\mathbb{E}}_t[y_{t+j}].$$

and where $y_t = \sum_g \pi_g y_{t,g}$. Note that $c_{g,t}$, the actual consumption of group g, may differ from $y_{g,t}$, the auxiliary variable introduced above. This will indeed be the case whenever $\overline{\mathbb{E}}_t[\mathcal{T}_{g,t}] \neq (1 - \chi\omega_g)s_{g,t}$ for some g and some t. Still, as long as (63) is satisfied, the economy-wide outcomes are determined in the manner described above—and coincide with those reported in the main text.

This completes the details behind Figure 4. Consider next what happens when condition (63) is violated and, as a result, wealth inequality can feed into the aggregate dynamics. In particular, impose $\mathcal{T}_{g,t} = 0$ for all g, t. If all groups had the same MPC, (63) and (64) would still hold; but then the heterogeneity in business-cycle exposure would also not matter. The interesting case is when fiscal policy is inactive and, in addition, there is joint heterogeneity in the business-cycle exposure and the MPC. This case is studied in Figure 5 in the main text.

F Application to Investment

A long tradition in macroeconomics that goes back to Hayashi (1982) and Abel and Blanchard (1983) has studied representative-agent models in which the firms face a cost in adjusting their capital stock. In this literature, the adjustment cost is specified as follows:

$$\text{Cost}_t = \Phi\left(\frac{I_t}{K_{t-1}}\right) \tag{65}$$

where I_t denotes the rate of investment, K_{t-1} denotes the capital stock inherited from the previous period, and Φ is a convex function. This specification gives the level of investment as a decreasing function of Tobin's Q. It also generates aggregate investment responses that are broadly in line with those predicted by more realistic, heterogeneous-agent models that account for the dynamics of investment at the firm or plant level (Caballero and Engel, 1999; Bachmann, Caballero, and Engel, 2013; Khan and Thomas, 2008).[41]

By contrast, the DSGE literature that follows Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2007) assumes that the firms face a cost in adjusting, not their capital stock, but rather their rate of investment. That is, this literature specifies the adjustment cost as follows:

$$\text{Cost}_t = \Psi\left(\frac{I_t}{I_{t-1}}\right) \tag{66}$$

As with the Hybrid NKPC, this specification was adopted because it allows the theory to generate sluggish aggregate investment responses to monetary and other shocks. But it has no obvious analogue in the literature that accounts for the dynamics of investment at the firm or plant level.

In the sequel, we set up a model of aggregate investment with two key features: first, the adjustment cost takes the form seen in condition (65); and second, the investments of different firms are strategic complements because of an aggregate demand externality. We then augment this model with incomplete information and show that it becomes observationally equivalent to a model in which the adjustment cost takes the form seen in condition (66). This illustrates how incomplete information can merge the gap between the different strands of the literature and help reconcile the dominant DSGE practice

[41] These works differ on the importance they attribute to heterogeneity, lumpiness, and non-linearities, but appear to share the prediction that the impulse response of aggregate investment is peaked on impact. They therefore do not provide a micro-foundation of the kind of sluggish investment dynamics featured in the DSGE literature.

with the relevant microeconomic evidence on investment.

Let us fill in the details. We consider an AK model with costs to adjusting the capital stock. There is a continuum of monopolistic competitive firms, indexed by i and producing different varieties of intermediate investment goods. The final investment good is a CES aggregator of intermediate investment goods. Letting $X_{i,t}$ denote the investment good produced by firm i, we have that the aggregate investment is given by

$$I_t = \left[\int X_{i,t}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}.$$

And letting $Q_{i,t}$ denote the price faced by firm i, we have that the investment price index is given by

$$Q_t = \left[\int Q_{i,t}^{1-\sigma} \right]^{\frac{1}{1-\sigma}}.$$

A representative final goods producer has perfect information and purchases investment goods to maximize its discounted profit

$$\max_{\{K_t, I_t\}} \sum_{t=0}^{\infty} \chi^t \mathbb{E}_0 \left[\exp(\xi_t) A K_t - Q_t I_t - \Phi\left(\frac{I_t}{K_t}\right) K_t \right],$$

subject to

$$K_{t+1} = K_t + I_t.$$

Here, the fundamental shock, ξ_t, is an exogenous productivity shock to the final goods production, and $\Phi\left(\frac{I_t}{K_t}\right) K_t$ represents the quadratic capital-adjustment cost. The following functional form is assumed:

$$\Phi\left(\frac{I_t}{K_t}\right) = \frac{1}{2}\psi \left(\frac{I_t}{K_t}\right)^2.$$

Let $Z_t \equiv \frac{I_t}{K_t}$ denote the investment-to-capital ratio. On a balanced growth path, this ratio and the price for the investment goods remain constant, i.e., $Z_t = Z$ and $Q_t = Q$. The log-linearized version of the final goods producer's optimal condition around the balanced growth path can be written as

$$Qq_t + \psi Z z_t = \chi \mathbb{E}_t \Big[A\xi_{t+1} + Qq_{t+1} + \psi Z(1+Z) z_{t+1} \Big]. \tag{67}$$

When the producers of the intermediate investment goods choose their production scale, they may not observe the underlying fundamental ξ_t perfectly. As a result, they have to make their decision based on their expectations about fundamentals and others' decisions. Letting

$$\max_{X_{i,t}} \mathbb{E}_{i,t} \left[Q_{i,t} X_{i,t} - c X_{i,t} \right],$$

subject to

$$Q_{i,t} = \left(\frac{X_{i,t}}{I_t}\right)^{-\frac{1}{\sigma}} Q_t.$$

Define $Z_{i,t} \equiv \frac{X_{i,t}}{K_t}$ as the firm-specific investment-to-capital ratio, and the log-linearized version of the optimal choice of $X_{i,t}$ is

$$z_{i,t} = \mathbb{E}_{i,t}\left[z_t + \sigma q_t\right].$$

In steady state, the price Q simply equals the markup over marginal cost c,

$$Q = \frac{\sigma}{\sigma - 1} c,$$

and the investment-to-capital ratio Z solves the quadratic equation

$$Q + \psi Z = \chi \left(A + Q + \psi Z + \psi Z^2 - \frac{1}{2}\psi Z^2 \right).$$

Frictionless Benchmark. If all intermediate firms observe ξ_t perfectly, then we have

$$z_{i,t} = z_t + \sigma q_t$$

Aggregation implies that $z_{i,t} = z_t$ and $q_t = 0$. It follows that z_t obeys the following Euler condition:

$$z_t = \varphi \xi_t + \delta \mathbb{E}_t\left[z_{t+1}\right]$$

where

$$\varphi = \frac{\rho \chi A}{\psi Z} \qquad \text{and} \qquad \delta = \chi(1 + Z).$$

Incomplete Information. Suppose now that firms receive a noisy signal about the fundamental ξ_t as in Section 2. Here, we make the same simplifying assumption as in the NKPC application. We assume that firms observe current z_t, but preclude them from extracting information from it. Together with the pricing equation (67), the aggregate investment dynamics follow

$$z_t = \frac{\rho \chi A}{\psi Z} \sum_{k=0}^{\infty} \chi^k \overline{\mathbb{E}}_t[\xi_{t+k}] + \chi Z \sum_{k=0}^{\infty} \chi^k \overline{\mathbb{E}}_t[z_{t+k+1}]$$

The investment dynamics can be understood as the solution to the dynamic beauty contest studied in Section 2 by letting

$$\varphi = \frac{\rho \chi A}{\psi Z}, \qquad \beta = \chi, \qquad \text{and} \qquad \gamma = \chi Z.$$

It is then immediate that when information is incomplete, there exist $\omega_f < 1$ and $\omega_b > 0$ such that the equilibrium process for investment solves the following equation:

$$z_t = \varphi \xi_t + \omega_f \delta \mathbb{E}_t[z_{t+1}] + \omega_b z_{t-1}.$$

Finally, it straightforward to show that the above equation is of the same type as the one that governs investment in a complete-information model where the adjustment cost is in terms of the investment rate, namely a model in which the final good producer's problem is modified as follows:

$$\max_{\{K_t, I_t\}} \sum_{t=0}^{\infty} \chi^t \mathbb{E}_0 \left[\exp(\xi_t) AK_t - Q_t I_t - \Psi\left(\frac{I_t}{\widetilde{I}_{t-1}}\right) I_t \right],$$

where $\widetilde{I}_t$ is the aggregate investment.

G Application to Asset Prices

Consider a log-linearized version of the standard asset-pricing condition in an infinite horizon, representative-agent model:

$$p_t = \mathbb{E}_t[d_{t+1}] + \chi \mathbb{E}_t[p_{t+1}],$$

where p_t is the price of the asset in period t, d_{t+1} is its dividend in the next period, $\mathbb{E}_t$ is the expectation of the representative agent, and χ is his discount factor. Iterating the above condition gives the equilibrium price as the expected present discounted value of the future dividends.

By assuming a representative agent, the above condition conceals the importance of higher-order beliefs. A number of works have sought to unearth that role by considering variants with heterogeneously informed, short-term traders, in the tradition of Singleton (1987); see, for example, Allen, Morris, and Shin (2006), Kasa, Walker, and Whiteman (2014), and Nimark (2017). We can capture these works in our setting by modifying the equilibrium pricing condition as follows:

$$p_t = \overline{\mathbb{E}}_t[d_{t+1}] + \chi \overline{\mathbb{E}}_t[p_{t+1}] + \epsilon_t,$$

where $\overline{\mathbb{E}}_t$ is the *average* expectation of the traders in period t and ϵ_t is an i.i.d shock interpreted as the price effect of noisy traders. The key idea embedded in the above condition is that, as long as the traders have different information and there are limits to arbitrage, asset markets are likely to behave like (dynamic) beauty contests.

Let us now assume that the dividend is given by $d_{t+1} = \xi_t + u_{t+1}$, where ξ_t follows an AR(1) process and u_{t+1} is i.i.d. over time, and that the information of the typical trader can be represented by a series of private signals as in condition (13).[42] Applying our results, and using the fact that $\xi_t = \mathbb{E}_t[d_{t+1}]$, we then have that the component of the equilibrium asset price that is driven by ξ_t obeys the following law of motion, for some $\omega_f < 1$ and $\omega_b > 0$:

$$p_t = \mathbb{E}_t[d_{t+1}] + \omega_f \chi \mathbb{E}_t[p_{t+1}] + \omega_b p_{t-1}, \tag{68}$$

[42] Here, we are abstracting from the complications of the endogenous revelation of information and we think of the signals in (13) as convenient proxies for all the information of the typical trader. One can also interpret this as a setting in which the dividend is observable (and hence so is the price, which is measurable in the dividend) and the assumed signals are the representation of a form of rational inattention. Last but not least, we have verified that the solution with endogenous information can be approximated very well by the solution obtained with exogenous information.

where $\mathbb{E}_t[\cdot]$ is the fully-information, rational expectations. We thus have that asset prices can display both myopia, in the form of $\omega_f < 1$, and momentum, or predictability, in the form of $\omega_b > 0$.

Although they do not contain such an observational-equivalence result, Kasa, Walker, and Whiteman (2014) have already pointed out that incomplete information and higher-order uncertainty can help explain momentum and predictability in asset prices. Our result offers a sharp illustration of this insight and blends it with the insight regarding myopia.

In the present context, the latter insight seems to challenge the asset-price literature that emphasizes long-run risks: news about the long-run fundamentals may be heavily discounted when there is higher-order uncertainty. Finally, our result suggests that both kinds of distortions are likely to be greater at the level of the entire stock market than at the level of the stock of a particular firm insofar as financial frictions and GE effects cause the trades to be strategic complements at the macro level even if they are strategic substitutes at the micro level, which in turn may help rationalize Samuelson's dictum (Jung and Shiller, 2005). We leave the exploration of these—admittedly speculative—ideas open for future research.

We conclude by iterating that the exact form of condition (68) relies on assuming away the role of the equilibrium price as an endogenous public signal. This may be an important omission for certain counterfactuals. But as indicated by the exercise conducted at the end of Section 6, the quantitative implications may be similar provided that the theory is disciplined with the relevant evidence on expectations.

H Robustness of Main Insights

Although our observational-equivalence result depends on stringent assumptions about the process of the fundamental and the available signals, it encapsulates a few broader insights, which in turn justify the perspective put forward in our paper.

The broader insights concerning the role of incomplete information and especially that of higher-order uncertainty can be traced in various previous works, including Angeletos and Lian (2018), Morris and Shin (2006), Nimark (2008), and Woodford (2003). But like our paper, these earlier work rely on strong assumptions about the underlying process of the fundamental, as well as about the information structure.

In this appendix, we relax completely the restrictions on the stochastic process for the fundamental. We then use a different, flexible but not entirely free, specification of the information structure to obtain a close-form characterization of the dynamics of the equilibrium outcome and the entire belief hierarchy. Our exact observational equivalence result is lost, but a generalization of the insights about myopia, anchoring and higher-order beliefs obtains.

Setup. We henceforth let the fundamental ξ_t follow a flexible, possibly infinite-order, MA process:

$$\xi_t = \sum_{k=0}^{\infty} \rho_k \eta_{t-k}, \tag{69}$$

where the sequence $\{\rho_k\}_{k=0}^{\infty}$ is non-negative and square summable. Clearly, the AR(1) process assumed earlier on is nested as a special case where $\rho_k = \rho^k$ for all $k \geq 0$. The present specification allows for richer, possibly hump-shaped, dynamics in the fundamental, as well as for "news shocks," that is, for innovations that shift the fundamental only after a delay.

Next, for every i and t, we let the incremental information received by agent i in period t be given by the series $\{x_{i,t,t-k}\}_{k=0}^{\infty}$, where

$$x_{i,t,t-k} = \eta_{t-k} + \epsilon_{i,t,t-k} \quad \forall k,$$

where $\epsilon_{i,t,t-k} \sim \mathcal{N}(0,(\tau_k)^{-2})$ is i.i.d. across i and t, uncorrelated across k, and orthogonal to the past, current, and future innovations in the fundamental, and where the sequence $\{\tau_k\}_{k=0}^{\infty}$ is non-negative and non-decreasing. In plain words, whereas our baseline specification has the agents observe a signal about the concurrent fundamental in each period, the new specification lets them observe a series of signals about the entire history of the underlying past and current innovations.

Although this specification may look exotic at first glance, it actually nest sticky information as a special case. We will verify this momentarily. It also preserves two key features of our baseline setting: it allows information to be incomplete at any given point of time; it lets more precise information and higher levels of common knowledge to be obtained as time passes.

Still, the present specification differs from our baseline one in two respects. First, it "orthogonalizes" the information structure in the sense that, for every t, every k, and every $k' \neq k$, the signals received at or prior to date t about the shock η_{t-k} are independent of the signals received about the shock $\eta_{t-k'}$. Second, it allows for more flexible learning dynamics in the sense that the precision τ_k does not have to be flat in k: the quality of the incremental information received in any given period about a past shock may either increase or decrease with the lag since the shock has occurred.

The first property is essential for tractability. The pertinent literature has struggled to solve for, or accurately approximate, the complex fixed point between the equilibrium dynamics and the Kalman filtering that obtains in dynamic models with incomplete information, especially in the presence of endogenous signals; see, for example, Nimark (2017). By adopting the aforementioned orthogonalization, we cut the Gordian knot and facilitate a closed-form solution of the entire dynamic structure of the higher-order beliefs and of the equilibrium outcome.[43] The second property then permits us, not only to ac-

[43] Such an orthogonalization may not square well with rational inattention or endogenous learning: in these contexts, the available signals may naturally confound information about current and past innovations, or even about entirely different kinds of fundamentals. The approach taken here is therefore, not a panacea, but rather a sharp instrument for understanding the specific friction we are after in this paper, namely the inertia of first- and higher-order beliefs. The possible confusion of different shocks is a conceptual distinct matter, outside the scope of this paper.

commodate a more flexible learning dynamics, but also to disentangle the speed of learning from level of noise—a disentangling that was not possible in Section 3 because a single parameter, σ, controlled both objects at once.

Dynamics of Higher-Order Beliefs. The information regarding η_{t-k} that an agent has accumulated up to, and including, period t can be represented by a sufficient statistic, given by

$$\tilde{x}_{i,t}^{k} = \sum_{j=0}^{k} \frac{\tau_j}{\pi_k} x_{i,t-j,t-k},$$

where $\pi_k \equiv \sum_{j=0}^{k} \tau_j$. That is, the sufficient statistic is constructed by taking a weighted average of all the available signals, with the weight of each signal being proportional to its precision; and the precision of the statistic is the sum of the precisions of the signals. Letting $\lambda_k \equiv \frac{\pi_k}{\sigma_\eta^{-2}+\pi_k}$, we have that $\mathbb{E}_{i,t}[\eta_{t-k}] = \lambda_k \tilde{x}_{i,t}^{k}$, which in turn implies $\overline{\mathbb{E}}_t[\eta_{t-k}] = \lambda_k \eta_{t-k}$ and therefore

$$\overline{\mathbb{E}}_t[\xi_t] = \overline{\mathbb{E}}_t\left[\sum_{k=0}^{\infty} \rho_k \eta_{t-k}\right] = \sum_{k=0}^{\infty} f_{1,k}\eta_{t-k}, \qquad \text{with} \qquad f_{1,k} = \lambda_k \rho_k. \tag{70}$$

The sequence $\mathbf{F}_1 \equiv \{f_{1,k}\}_{k=0}^{\infty} = \{\lambda_k \rho_k\}_{k=0}^{\infty}$ identifies the IRF of the average first-order forecast to an innovation. By comparison, the IRF of the fundamental itself is given by the sequence $\{\rho_k\}_{k=0}^{\infty}$. It follows that the relation of the two IRFs is pinned down by the sequence $\{\lambda_k\}_{k=0}^{\infty}$, which describes the dynamics of learning. In particular, the smaller λ_0 is (i.e., the less precise the initial information is), the larger the initial initial gap between the two IRFs (i.e., a larger the initial forecast error). And the slower λ_k increases with k (i.e., the slower the learning over time), the longer it takes for that gap (and the average forecast) to disappear.

These properties are intuitive and are shared by the specification studied in the rest of the paper. In the information structure specified in Section 3, the initial precision is tied with the subsequent speed of learning. By contrast, the present specification disentangles the two. As shown next, it also allows for a simple characterization of the IRFs of the higher-order beliefs, which is what we are after.

Consider first the forward-looking higher-order beliefs. Applying condition (70) to period $t+1$ and taking the period-t average expectation, we get

$$\overline{\mathbb{F}}_t^2[\xi_{t+1}] \equiv \overline{\mathbb{E}}_t\left[\overline{\mathbb{E}}_{t+1}[\xi_{t+1}]\right] = \overline{\mathbb{E}}_t\left[\sum_{k=0}^{\infty} \lambda_k \rho_k \eta_{t+1-k}\right] = \sum_{k=0}^{\infty} \lambda_k \lambda_{k+1} \rho_{k+1} \eta_{t-k}$$

Notice here, agents in period t understand that in period $t+1$ the average forecast will be improved, and this is why λ_{k+1} shows up in the expression. By induction, for all $h \geq 2$, the h-th order, forward-looking belief is given by

$$\overline{\mathbb{F}}_t^h[\xi_{t+h-1}] = \sum_{k=0}^{\infty} f_{h,k}\eta_{t-k}, \qquad \text{with} \qquad f_{h,k} = \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} \rho_{k+h-1}. \tag{71}$$

The increasing components in the product $\lambda_k \lambda_{k+1} ... \lambda_{k+h-1}$ seen above capture the anticipation of learning. We revisit this point at the end of this section.

The set of sequences $\mathbf{F}_h = \{f_{h,k}\}_{k=0}^{\infty}$, for $h \geq 2$, provides a complete characterization of the IRFs of the relevant, forward-looking, higher-order beliefs. Note that $\frac{\partial \mathbb{E}[\xi_{t+h}|\eta_{t-k}]}{\partial \eta_{t-k}} = \rho_{k+h-1}$. It follows that the ratio $\frac{f_{h,k}}{\rho_{k+h-1}}$ measures the effect of an innovation on the relevant h-th order belief relative to its effect on the fundamental. When information is complete, this ratio is identically 1 for all k and h. When, instead, information is incomplete, this ratio is given by

$$\frac{f_{h,k}}{\rho_{k+h-1}} = \lambda_k \lambda_{k+1} ... \lambda_{k+h-1}.$$

The following result is thus immediate.

Proposition 13. *Consider the ratio $\frac{f_{h,k}}{\rho_{k+h-1}}$, which measures the effect at lag k of an innovation on the h-th order forward-looking belief relative to its effect on the fundamental.*

(i) For all k and all h, this ratio is strictly between 0 and 1.

(ii) For any k, this is decreasing in h.

(iii) For any h, this ratio is increasing in k.

(iv) As $k \to \infty$, this ratio converges to 1 for any $h \geq 2$ if and only if it converges for $h = 1$, and this in turn is true if and only if $\lambda_k \to 1$.

These properties shed light on the dynamic structure of higher-order beliefs. Part (i) states that, for any belief order h and any lag k, the impact of a shock on the h-th order belief is lower than that on the fundamental itself. Part (ii) states that higher-order beliefs move less than lower-order beliefs both on impact and at any lag. Part (iii) states that that the gap between the belief of any order and the fundamental decreases as the lag increases; this captures the effect of learning. Part (iv) states that, regardless of h, the gap vanishes in the limit as $k \to \infty$ if and only if $\lambda_k \to 1$, that is, if and only if the learning is bounded away from zero.

Sticky information. We now verify the claim made in the main text that the assumed information structure nests sticky information Ãă la Mankiw and Reis (2002).

Each agent updates her information set with probability $1 - q \in (0, 1)$ in each period. When she updates, she gets to see the entire state of Nature. Otherwise, her information remains the same as in the previous period.

Consider now an arbitrary innovation η_t in some period t. A fraction $1-q$ of the population becomes aware of it immediately and hence $\overline{\mathbb{E}}_t[\eta_t] = (1-q)\eta_t$. A period later, an additional $(1-q)q$ fraction becomes aware of it and hence $\overline{\mathbb{E}}_{t+1}[\eta_t] = (1-q^2)\eta_t$. And so on. It follows that sticky information Ãă la Mankiw and Reis (2002) is nested in the present setting under the following restriction on the sequence $\{\lambda_k\}$:

$$\lambda_k = 1 - q^k.$$

Furthermore, under this interpretation, endogenizing the frequency $1-q$ with which agents update their information maps merely to endogenizing the sequence $\{\lambda_\tau\}_{\tau=0}^{\infty}$. Conditional on it, all the results presented in the sequel remain intact. This hints to the possible robustness of our insights to endogenous information acquisition, an issue that we however abstract from: in what follows, we treat $\{\lambda_\tau\}_{\tau=0}^{\infty}$ as exogenous.

Myopia and Anchoring. To see how these properties drive the equilibrium behavior, we henceforth restrict $\beta = 0$ and normalize $\varphi = 1$. As noted earlier, the law of motion for the equilibrium outcome is then given by $a_t = \overline{\mathbb{E}}_t[\xi_t] + \gamma\overline{\mathbb{E}}_t[a_{t+1}]$, which in turn implies that $a_t = \sum_{h=1}^{\infty} \gamma^{h-1} \overline{\mathbb{F}}_t^h [\xi_{t+h-1}]$. From the preceding characterization of the higher-order beliefs $\overline{\mathbb{F}}_t^h [\xi_{t+h-1}]$, it follows that

$$a_t = \sum_{k=0}^{\infty} g_k \eta_{t-k}, \quad \text{with} \quad g_k = \sum_{h=1}^{\infty} \gamma^{h-1} f_{h,k} = \left\{ \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} \rho_{k+h-1} \right\}. \tag{72}$$

This makes clear how the IRF of the equilibrium outcome is connected to the IRFs of the first- and higher-order beliefs. Importantly, the higher γ is, the more the dynamics of the equilibrium outcome tracks the dynamics higher-order beliefs relative to the dynamics of lower-order beliefs. On the other hand, when the growth rate of the IRF of the fundamental $\frac{\rho_{k+1}}{\rho_k}$ is higher, it also increases the relative importance of higher-order beliefs.[44]

We are now ready to explain our result regarding myopia. For this purpose, it is best to abstract from learning and focus on how the mere presence of higher-order uncertainty affects the beliefs about the future. In the absence of learning, $\lambda_k = \lambda$ for all k and for some $\lambda \in (0,1)$. The aforementioned formula for the IRF coefficients then reduces to the following:

$$g_k = \left\{ \sum_{h=1}^{\infty} (\gamma\lambda)^{h-1} \rho_{k+h-1} \right\} \lambda.$$

Clearly, this the same IRF as that of a complete-information, representative-economy economy in which the equilibrium dynamics satisfy

$$a_t = \xi_t' + \gamma' \mathbb{E}_t[a_{t+1}], \tag{73}$$

where $\xi_t' \equiv \lambda \xi_t$ and $\gamma' \equiv \gamma\lambda$. It is therefore as if the fundamental is less volatile and, in addition, the agents are less forward-looking. The first effect stems from first-order uncertainty: it is present simply because the forecast of the fundamental move less than one-to-one with the true fundamental. The second effect originates in higher-order uncertainty: it is present because the forecasts of the actions of others move *even* less than the forecast of the fundamental.

[44]The last point is particularly clear if we set $\rho_k = \rho^k$ (meaning that ξ_t follows an AR(1) process). In this case, the initial response is given by

$$g_0 = \sum_{h=1}^{\infty} (\gamma\rho)^{h-1} \lambda_0 \lambda_1 \dots \lambda_{h-1},$$

from which it is evident that the importance of higher-order beliefs increases with both γ and ρ. This further illustrates the point made in Section 3.4 regarding the role of the persistence of the fundamental.

This is the crux of the forward-looking component of our observational-equivalence result (that is, the one regarding myopia). Note in particular that the extra discounting of the future remains present even if when if control for the impact of the informational friction on first-order beliefs. Indeed, replacing ξ'_t with ξ_t in the above shuts down the effect of first-order uncertainty. And yet, the extra discounting survives, reflecting the role of higher-order uncertainty. This complements the related points we make in Section 3.5.

So far, we shed light on the source of myopia, while shutting down the role of learning. We next elaborate on the robustness of the above insights to the presence of learning and, most importantly, on how the presence of learning and its interaction with higher-order uncertainty drive the backward-looking component of our observational-equivalence result.

To this goal, and as a benchmark for comparison, we consider a variant economy in which all agents share the same subjective belief about ξ_t, this belief happens to coincide with the average first-order belief in the original economy, and these facts are common knowledge. The equilibrium outcome in this economy is proportional to the subjective belief of ξ_t and is given by

$$a_t = \sum_{k=0}^{\infty} \widehat{g}_k \eta_{t-k}, \qquad \text{with} \qquad \widehat{g}_k = \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k \rho_{k+h-1}.$$

This resembles the complete-information benchmark in that the outcome is pined down by the first-order belief of ξ_t, but allows this belief to adjust sluggishly to the underlying innovations in ξ_t.

By construction, the variant economy preserves the effects of learning on first-order beliefs but shuts down the interaction of learning with higher-order uncertainty. It follows that the comparison of this economy with the original economy reveals the role of this interaction.

Proposition 14. *Let $\{g_k\}$ and $\{\widehat{g}_k\}$ denote the Impulse Response Function of the equilibrium outcome in the two economies described above.*

(i) $0 < g_k < \widehat{g}_k$ *for all* $k \geq 0$

(ii) If $\frac{\rho_k}{\rho_{k-1}} \geq \frac{\rho_{k+1}}{\rho_k}$ *and* $\rho_k > 0$ *for all* $k > 0$*, then* $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$ *for all* $k \geq 0$

Consider property (i), in particular the property that $g_k < \widehat{g}_k$. This property means that our economy exhibits a uniformly smaller dynamic response for the equilibrium outcome than the aforementioned economy, in which higher-order uncertainty is shut down. But note that the two economies share the following law of motion:

$$a_t = \varphi \overline{\mathbb{E}}_t[\xi_t] + \gamma \overline{\mathbb{E}}_t[a_{t+1}]. \tag{74}$$

Furthermore, the two economies share the same dynamic response for $\overline{\mathbb{E}}_t[\xi_t]$. It follows that the response for a_t in our economy is smaller than that of the variant economy because, and only because, the response of $\overline{\mathbb{E}}_t[a_{t+1}]$ is also smaller in our economy. This verifies that the precise role of higher-order uncertainty is to arrest the response of the expectations of the future outcome (the future actions of others) beyond and above how much the first-order uncertainty (the unobservability of ξ_t) arrests the response of the expectations of the future fundamental.

A complementary way of seeing this point is to note that g_k satisfies the following recursion:

$$g_k = f_{1,k} + \lambda_k \gamma g_{k+1}. \tag{75}$$

The first term in the right-hand side of this recursion corresponds to the average expectation of the future fundamental. The second term corresponds the average expectation of the future outcome (the actions of others). The role of first-order uncertainty is captured by the fact that $f_{1,k}$ is lower than ρ_k. The role of higher-order uncertainty is captured by the presence of λ_k in the second term: it is as if the discount factor γ has been replaced by a discount factor equal to $\lambda_k\gamma$, which is strictly less than γ. This represents a generalization of the form of myopia seen in condition (73). There, learning was shut down, so that that λ_k and the extra discounting of the future were invariant in the horizon k. Here, the additional discounting varies with the horizon because of the anticipation of future learning (namely, the knowledge that λ_k will increase with k).

Consider next property (ii), namely the property that

$$\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}.$$

This property helps explain the backward-looking component of our observational-equivalence result (that is, the one regarding anchoring).

To start with, consider the variant economy, in which higher-order uncertainty is shut down. The impact of a shock $k+1$ periods from now relative to its impact k periods from now is given by

$$\frac{\widehat{g}_{k+1}}{\widehat{g}_k} = \frac{\lambda_{k+1}}{\lambda_k}\frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}} > \frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}}.$$

The inequality captures the effect of learning on first-order beliefs. Had information being perfect, we would have had $\frac{\widehat{g}_{k+1}}{\widehat{g}_k} = \frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}}$; now, we instead have $\frac{\widehat{g}_{k+1}}{\widehat{g}_k} > \frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}}$. This means that, in the variant economy, the impact of the shock on the equilibrium outcome can build force over time because, and only because, learning allows for a gradual build up in first-order beliefs.[45]

Consider now our economy, in which higher-order uncertainty is present. We now have

$$\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$$

This means that higher-order uncertainty amplifies the build-up effect of learning: as time passes, the impact of the shock on the equilibrium outcome builds force more rapidly in our economy than in the

[45]This is easiest to see when $\rho_k = 1$ (i.e., the fundamental follows a random walk), for then $\widehat{g}_{k+1}$ is necessarily higher than $\widehat{g}_k$ for all k. In the AR(1) case where $\rho_k = \rho^k$ with $\rho < 1$, $\widehat{g}_{k+1}$ can be either higher or lower than $\widehat{g}_k$, depending on the balance between two opposing forces: the build-up effect of learning and the mean-reversion in the fundamental.

variant economy. But since the impact is always lower in our economy,[46] this means that the IRF of the equilibrium outcome is likely to display a more pronounced hump shape in our economy than in the variant economy. Indeed, the following is a directly corollary of the above property.

Corollary 5. *Let the variant economy display a hump-shaped response: $\{\widehat{g}_k\}$ is single peaked at $k = k^b$ for some $k^b \geq 1$. Then, the equilibrium outcome also displays a hump-shaped response: $\{g_k\}$ is also single peaked at $k = k^g$. Furthermore, the peak of the equilibrium response is after the peak of the variant economy: $k^g \geq k^b$ necessarily, and $k^g > k^b$ for an open set of $\{\lambda_k\}$ sequences.*

To interpret this result, think of k as a continuous variable and, similarly, think of λ_k, $\widehat{g}_k$, and g_k as differentiable functions of k. If $\widehat{g}_k$ is hump-shaped with a peak at $k = k_b > 0$, it must be that $\widehat{g}_k$ is weakly increasing prior to k_b and locally flat at k_b. But since we have proved that the growth rate of g_k is strictly higher than that of $\widehat{g}_k$, this means that g_k attains its maximum at a point k_g that is strictly above k^b. In the result stated above, the logic is the same. The only twist is that, because k is discrete, we must either relax $k_g > k_b$ to $k_g \geq k_b$ or put restrictions on $\{\lambda_k\}$ so as to guarantee that $k_g \geq k_b + 1$.

Summing up, learning by itself contributes towards a gradual build up of the impact of any given shock on the equilibrium outcome; but its interaction with higher-order uncertainty makes this build up even more pronounced. It is precisely these properties that are encapsulated in the backward-looking component of our observational equivalence result: the coefficient ω_b, which captures the endogenous build up in the equilibrium dynamics, is positive because of learning and it is higher the higher the importance of higher-order uncertainty.

Multiple Fundamental Shocks. So far, we have focused on the case where there is a single fundamental shock. Now we extend the analysis to a case where multiple fundamental shocks are present. On one hand, we will show that relative to the frictionless benchmark, when these shocks cannot be perfectly separated, agents may overact to some of these shocks and underact to the others when we focus on the PE effects, as in Lucas (1976). On the other hand, we will show that higher-order uncertainty, which exclusively related to the GE effects, still results in distortions in the form of myopia and anchoring *relative* to its complete-information counterpart.

Suppose that the best response is

$$a_{i,t} = \mathbb{E}_{i,t}[\phi_1 \xi_t^1 + \phi_2 \xi_t^2] + \gamma \mathbb{E}_{i,t}[a_{t+1}],$$

where the two fundamental shocks are driven by two different innovations η_t^1 and η_t^2

$$\xi_t^1 = \sum_{k=0}^{\infty} \rho_k^1 \eta_{t-k}^1, \qquad \text{and} \qquad \xi_t^2 = \sum_{k=0}^{\infty} \rho_k^2 \eta_{t-k}^2.$$

We assume that agents do not observe separate signals about the innovations to the two fundamental

[46] Recall, this is by property (i) of Proposition 14.

shocks, but only a sum of them, i.e.,

$$x_{i,t,t-k} = \eta^1_{t-k} + \eta^2_{t-k} + \epsilon_{i,t,t-k} \quad \forall k.$$

This signal structure is the same as before if agents only care about the sum $\eta_t \equiv \eta^1_t + \eta^2_t$, and it follows that

$$\overline{\mathbb{E}}_t[\eta_{t-k}] = \lambda_k.$$

where the sequence of λ_k is defined in a similar way as before. The average expectations on each of the aggregate innovations is given by

$$\overline{\mathbb{E}}_t[\eta^1_{t-k}] = \varpi_1 \lambda_k, \qquad \text{and} \qquad \overline{\mathbb{E}}_t[\eta^2_{t-k}] = \varpi_2 \lambda_k,$$

where the weights ϕ_1 and ϕ_2 depend on the relative volatility of η^1_t versus η^2_t, satisfying $\varpi_1 + \varpi_2 = 1$.

First consider the case where $\gamma = 0$, that is, only the PE consideration is at work. The average expectations about the fundamental are given by

$$\overline{\mathbb{E}}_t\left[\phi_1 \xi^1_t\right] = \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho^1_k \eta_{t-k} = \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho^1_k \eta^1_{t-k} + \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho^1_k \eta^2_{t-k},$$

$$\overline{\mathbb{E}}_t\left[\phi_2 \xi^1_t\right] = \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho^2_k \eta_{t-k} = \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho^2_k \eta^1_{t-k} + \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho^2_k \eta^2_{t-k}.$$

In the absence of GE consideration and higher-order expectation, we can see that agents may overact to some of the fundamental. Consider the response to innovation of the first fundamental, η^1_t. In the frictionless case, $\overline{\mathbb{E}}_t\left[\omega_1 \xi^1_t\right] = \omega_1 \sum_{k=0}^{\infty} \rho^1_k \eta^1_{t-k}$. The average expectation of ξ^1_t under incomplete information is modified in two ways: on one hand, it is attenuated by the terms $\{\lambda_k \phi_1\}$; on the other hand, it also responds to η^2_t due to informational frictions. The total effects could well be a higher response overall.

Now we turn to the effects of the GE consideration and higher-order uncertainty with $\gamma > 0$. The average higher-order expectations are given by

$$\overline{\mathbb{F}}^h_t\left[\omega_1 \xi^1_{t+h-1} + \omega_2 \xi^2_{t+h-1}\right] = \sum_{k=0}^{\infty} f_{h,k} \eta_{t-k}, \qquad \text{with} \qquad f_{h,k} = \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} (\omega_1 \phi_1 \rho^1_{k+h-1} + \omega_2 \phi_2 \rho^2_{k+h-1}).$$

Here, we utilize the property that agents cannot separate η^1_t from η^2_t and the expectations can be effectively written as functions of η_t.

Similar to the single-shock economy, the aggregate outcome can be written as

$$a_t = \sum_{k=0}^{\infty} g_k \eta_{t-k}, \quad \text{with} \quad g_k = \sum_{h=1}^{\infty} \gamma^{h-1} f_{h,k} = \left\{ \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} (\omega_1 \phi_1 \rho^1_{k+h-1} + \omega_2 \phi_2 \rho^2_{k+h-1}) \right\}. \tag{76}$$

In contrast, with complete but imperfect information that shares the same first-order belief, the aggregate outcome is

$$a_t = \sum_{k=0}^{\infty} \widehat{g}_k \eta_{t-k}, \quad \text{with} \quad \widehat{g}_k = \left\{ \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k (\omega_1 \phi_1 \rho^1_{k+h-1} + \omega_2 \phi_2 \rho^2_{k+h-1}) \right\}. \tag{77}$$

Define $\widehat{\xi}_t$ as

$$\widehat{\xi}_t \equiv \sum_{k=0}^{\infty} (\omega_1 \phi_1 \rho^1_k + \omega_2 \phi_2 \rho^2_k) \eta_{t-k}.$$

By replacing ξ_t by $\widehat{\xi}_t$, the analysis on myopia and anchoring in Proposition 14 extends to the current setting. Therefore, relative to the complete-information counterpart, the effects of additional myopia and anchoring remain the same when there exist multiple fundamental shocks.

Two Forms of Bounded Rationality. We now shed light on two additional points, which were anticipated earlier on: the role played by the anticipation that others will learn in the future; and the possible interaction of incomplete information with Level-k Thinking.

To illustrate the first point, we consider a behavioral variant where agents fail to anticipate that others will learn in the future. To simplify, we also set $\beta = 0$. Recall from equation (71), when agents are rational, the forward higher-order beliefs are

$$\overline{\mathbb{F}}^h_t [\xi_{t+h-1}] = \sum_{k=0}^{\infty} \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} \rho_{k+h-1} \eta_{t-k}.$$

In the variant economy, by shutting down the anticipation of learning, the nature of higher-order beliefs changes, as $\mathbb{E}_{i,t}\left[\overline{\mathbb{E}}_{t+k}\left[\xi_{t+q}\right]\right] = \mathbb{E}_{i,t}\left[\overline{\mathbb{E}}_t\left[\xi_{t+q}\right]\right]$ for $k, q \geq 0$, and the counterpart of $\overline{\mathbb{F}}^h_t[\xi_{t+h-1}]$ becomes

$$\overline{\mathbb{E}}^h_t [\xi_{t+h-1}] \equiv \overline{\mathbb{E}}_t \left[\overline{\mathbb{E}}_t [\ldots \overline{\mathbb{E}}_t [[\xi_{t+h-1}] \ldots]\right] = \sum_{k=0}^{\infty} \lambda^h_k \rho_{t+h-1} \eta_{t-k}.$$

Learning implies $\lambda_{k+1} > \lambda_k$, and the anticipation of learning implies $\lambda_k \lambda_{k+1} ... \lambda_{k+h-1} > \lambda^h_k$. As a result, higher-order beliefs in the behavioral variant under consideration vary *less* than those under rational expectations. By the same token, the aggregate outcome in this economy, which is given

$$a_t = \sum_{h=1}^{\infty} \gamma^{h-1} \overline{\mathbb{E}}^h_t [\xi_{t+h-1}],$$

behaves as if the myopia and anchoring are stronger than in the rational-expectations counterpart. In line with these observations, it can be shown that, if we go back to our baseline specification and impose that agents fail to anticipate that others will learn in the future, Proposition 3 continues to hold with the following modification: ω_f is smaller and ω_b is higher.

To illustrate the second point, we consider a variant that lets agents have limited depth of reasoning in the sense of Level-k Thinking. With level-0 thinking, agents believe that the aggregate outcome is fixed at zero for all t, but still form rational beliefs about the fundamental. Therefore, $a^0_{i,t} = \mathbb{E}_{i,t}[\xi_t]$, and the implied aggregate outcome for level-0 thinking is $a^0_t = \overline{\mathbb{E}}_t[\xi_t]$.

With level-1 thinking, agent i's action changes to

$$a_{i,t}^1 = \mathbb{E}_{i,t}[\xi_t] + \gamma \mathbb{E}_{i,t}[a_{t+1}^0] = \mathbb{E}_{i,t}[\xi_t] + \gamma \mathbb{E}_{i,t}\left[\overline{\mathbb{E}}_{t+1}[\xi_{t+1}]\right],$$

where the second-order higher-order belief shows up. By induction, the level-k outcome is given by

$$a_t^k = \sum_{h=1}^{k+1} \gamma^{h-1} \overline{\mathbb{F}}_t^h [\xi_{t+h-1}].$$

In a nutshell, Level-k Thinking truncates the hierarchy of beliefs at a finite order.

Compared with the rational-expectations economy that has been the focus of our analysis, the GE feedback effects in both of the aforementioned two variants are attenuated, and the resulting as-if myopia is strengthened. Furthermore, by selecting the depth of thinking, we can make sure that the second variant produces a similar degree of myopia as the first one.[47] That said, the source of the additional myopia is different. In the first, the relevant forward-looking higher-order beliefs have been replaced by myopic counterparts, which move less. In the second, the right, forward-looking higher-order beliefs are still at work, but they have been truncated at a finite point.

I Multiple Shocks

Our baseline specification has assumed that there is a single shock that drives the fundamental. In this section, we extend our analysis in the direction of Kohlhas and Walther (2019) to include both procyclical and countercyclical components, and show that a modified version of our main result holds.

Consider the following best response, which is similar to our baseline specification:

$$y_{it} = \varphi \mathbb{E}_{it}[\zeta_t] + \beta \mathbb{E}_{it}[y_{it+1}] + \gamma \mathbb{E}_{it}[y_{t+1}]. \tag{78}$$

But now allow the fundamental ζ_t to be driven by N different components:

$$\zeta_t = \sum_{j=1}^{N} d_{jt}, \quad \text{with} \quad d_{jt} = \kappa_j \xi_t + \epsilon_{j,t}.$$

The common shock among different components, ξ_t, follows an AR(1) process:

$$\xi_t = \rho \xi_{t-1} + \eta_t.$$

The component-specific shocks $\epsilon_{j,t} \sim \mathcal{N}(0, \tau_j^{-1})$ are i.i.d. across both j and t. The loading of component j on ξ_t is κ_j, which could be both positive or negative, capturing for procyclical or counter-cyclical components. Finally, $\sum_j \kappa_j = 1$.

In terms of the information structure, assume that each agent receives N private signals, one per

[47] This follows directly from the fact that impact of effect of an innovation in the first variant is bounded between those of the level-0 and the level-∞ outcome in the second variant.

component:

$$x_{i,j,t} = d_{jt} + u_{i,j,t}, \quad u_{i,j,t} \sim \mathcal{N}(0, \omega_j^{-1}).$$

This is the same structure considered in Kohlhas and Walther (2019), which leads to asymmetric attention by allowing heterogeneity in ω_j.

To see how this structure connects with our equivalence result, we turn to the following auxiliary best response in which only the persistent shock ξ_t is pay-off relevant:

$$a_{it} = \varphi \mathbb{E}_{it}[\xi_t] + \beta \mathbb{E}_{it}[a_{it+1}] + \gamma \mathbb{E}_{it}[a_{t+1}]. \tag{79}$$

This best response is exactly the same as our model. The aggregate outcome y_t from condition (78) is related to the aggregate outcome a_t from condition (79) in the following way:

$$\begin{aligned} y_t &= \varphi \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t[\zeta_{t+k}] + \gamma \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t[y_{t+k+1}] \\ &= \varphi \sum_{j=1}^{N} \overline{\mathbb{E}}_t[\epsilon_{j,t}] + \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t[\xi_{t+k}] + \gamma \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t[y_{t+k+1}] \\ &= \varphi \sum_{j=1}^{N} \overline{\mathbb{E}}_t[\epsilon_{j,t}] + a_t, \end{aligned}$$

where the last equality is due to that only the persistent shock ξ_t matters for y_{t+k} in the future, and the forecasts of the transitory shocks $\epsilon_{j,t}$ are zero. We conclude that

$$y_t = a_t + u_t,$$

where $u_t \equiv \varphi \sum_{j=1}^{N} \overline{\mathbb{E}}_t[\epsilon_{j,t}]$.

Consider how u_t is determined. To this goal, let us first compute the forecast of the persistent shock ξ_t. Since this object only involves the first-order belief, it is more convenient to consolidate the N different signals into a single one

$$x_{i,t} = \xi_t + \frac{1}{\tau} \sum_{j=1}^{N} \kappa_j (\tau_j^{-1} + \omega_j^{-1})^{-1} (\epsilon_{j,t} + u_{i,j,t}) \equiv \xi_t + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}(0, \tau^{-1}),$$

where $\tau = \sum_{j=1}^{N} \kappa_j^2 (\tau_j^{-1} + \omega_j^{-1})^{-1}$. That is, it is as if each agent observes a single signal, which however contains both idiosyncratic and aggregate noise—a hybrid of the private and public signals considered in Appendix B. Using this observation, we can compute the average forecast as follows:

$$\overline{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{g}{\rho}\right) \frac{1}{1 - gL} \sum_{j=1}^{N} \frac{\kappa_j^2 (\tau_j^{-1} + \omega_j^{-1})^{-1}}{\tau} (\xi_t + \kappa_j^{-1} \epsilon_{j,t}),$$

or equivalently

$$\overline{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{g}{\rho}\right) \frac{1}{1 - gL} (\xi_t + \epsilon_t), \tag{80}$$

where $g \equiv \frac{1}{2}\left[\rho + \frac{1}{\rho}(1+\tau) - \sqrt{\left(\rho + \frac{1}{\rho}(1+\tau)\right)^2 - 4}\right] < \rho$ and $\epsilon_t \equiv \frac{1}{\tau}\sum_{j=1}^N \kappa_j(\tau_j^{-1} + \omega_j^{-1})^{-1}\epsilon_{j,t}$. Next, denote with $\lambda_j \equiv \frac{\omega_j}{\tau_j + \omega_j} \in (0,1)$ the signal-to-noise ratio applied when inferring $\epsilon_{j,t}$ from $u_{i,j,t}$. The average forecast of the sum of component-specific shocks is given by

$$\sum_{j=1}^N \overline{\mathbb{E}}_t[\epsilon_{j,t}] = \sum_{j=1}^N \lambda_j(\kappa_j \xi_t + \epsilon_{j,t} - \kappa_j \overline{\mathbb{E}}_t[\xi_t]) = \sum_{j=1}^N \lambda_j \kappa_j(\xi_t - \overline{\mathbb{E}}_t[\xi_t]) + \sum_{j=1}^N \lambda_j \epsilon_{jt}. \tag{81}$$

It follows that the determination of u_t boils down to a pure forecasting problem spelled out by equations (80) and (81).

Consider next the determination of a_t. As already mentioned, this obtains from the same best responses as our model, with ξ_t been the sole fundamental. The information structure about it is more complicated that in our baseline analysis, as agents observed signals contaminated with both idiosyncratic and common noise. But a result similar to Proposition 12 in Appendix B applies. That is,

$$a_t = a_t^\xi + \nu_t,$$

where a_t^ξ, the fundamental component, obeys our observational equivalence result and v_t, the residual, is an AR(1) driven by the "noise" (here, the combination of the $\epsilon_{j,t}$'s). The only subtle difference is in the precise cubic that pins down ϑ (and thereby ω_f and ω_b).

To complete the picture, consider the projection of y_t on the history of ξ_t. This is given by

$$y_t^\xi = \frac{\widetilde{\varphi}}{1-gL}\eta_t + a_t^\xi,$$

where η_t is the innovation in ξ_t and $\widetilde{\varphi} \equiv \varphi \sum_{j=1}^N \lambda_j \kappa_j \frac{g}{\rho}$. We thus have that the IRF of y_t with respect to η_t is the sum of the AR(2) corresponding to $a_t^x i$ and of the AR(1) given by the first term above. Clearly, this term does not contribute to a hump-shape. Furthermore, it is likely to be quantitatively less important than a_t^ξ for the following reason: a_t^ξ consists of all the PE and GE effects across all the horizons, while $\frac{\widetilde{\varphi}}{1-gL}\eta_t$ captures only a fraction of the total PE effects. For instance, as explained in Appendix D.2, in our inflation applications GE effects are about 7 times as large as PE effects. This suggests that, in that context, the a_t^ξ term would easily overwhelm the other term.

Let us conclude with the following comment. Kohlhas and Walther (2019) have used a model of the type described above to show that asymmetric attention allocation to various components of the outcome may help reconcile the form of belief over-reaction documented in their paper with the form of belief under-reaction documented in CG. Our network extension in Section 8 allows one to consider a multi-sector economy in which different sectors have different exposures to the aggregate shock, either directly or indirectly via differential GE effects. This may provide a more detailed micro-foundation for pro- and counter-cyclical components of economic activity, along the lines suggested by the aforementioned paper. And it could help study the role of asymmetries in GE feedbacks, similarly in spirit to what we do in our HANK application in Section 7.[48]

J Additional Proofs

Proof of Lemma 1, Lemma 2, and Proposition 8

The demand schedule faced by an individual firm i in market m is given by

$$Y_{i,m,t} = \left(\frac{P_{i,m,t}}{P_{m,t}}\right)^{-\psi} \left(\frac{P_{m,t}}{P_t}\right)^{-\varepsilon} Y_t,$$

where ψ and ε are within- and across-market elasticities of substitution, respectively. The price index in market m and the aggregate price index are defined as

$$P_{m,t} = \left(\frac{1}{N}\sum_j P_{j,m,t}^{1-\psi}\right)^{\frac{1}{1-\psi}}, \qquad P_t = \left[\int_m P_{m,t}^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}}.$$

In the absence of nominal rigidity and informational frictions, an individual firm i in market m sets its price to maximize its profit in the current period

$$\max_{P_{i,m,t}} P_{i,m,t} Y_{i,m,t} - P_t C(Y_t) Y_{i,m,t},$$

where $C(Y_t)$ is the marginal real cost which depends on the aggregate economic condition. Using the following properties

$$\frac{\partial Y_{i,m,t}}{\partial P_{i,m,t}} = -\psi \frac{Y_{i,m,t}}{P_{i,m,t}} + (\psi - \varepsilon)\frac{Y_{i,m,t}}{P_{m,t}}\frac{\partial P_{m,t}}{\partial P_{i,m,t}}, \quad \text{and} \quad \frac{\partial P_{m,t}}{\partial P_{i,m,t}} = \frac{1}{N} P_{m,t}^{\psi} P_{i,m,t}^{-\psi},$$

the first-order condition is

$$(1-\psi)\frac{P_{i,m,t}}{P_t} + \psi C(Y_t) = \frac{\varepsilon - \psi}{N}\left(\frac{P_{i,m,t}}{P_t} - C(Y_t)\right)\left(\frac{P_{i,m,t}}{P_{m,t}}\right)^{1-\psi} = 0.$$

We assume that $C(Y_t) = C\exp(\text{mc}_t)$ where mc_t follows an AR(1) process

$$\text{mc}_t = \rho \text{mc}_{t-1} + \eta_t.$$

In steady state where $\text{mc}_t = 0$ and $P_{i,m,t} = P_{m,t} = P_t$, it follows that

$$C = \frac{\psi - 1 + \frac{\varepsilon - \psi}{N}}{\psi + \frac{\varepsilon-\psi}{N}}.$$

The log-linearized version of the first-order condition is

$$(1-\psi)(p_{i,m,t} - p_t) + \psi C \text{mc}_t = \frac{\varepsilon - \psi}{N}\Big(p_{i,m,t} - p_t - C\text{mc}_t + (1-C)(1-\psi)(p_{i,m,t} - p_{m,t})\Big),$$

[48] There, asymmetric GE feedbacks emerge because of heterogeneous MPCs and heterogeneous exposures of income to business-cycle fluctuations.

which leads to the following best response

$$p_{i,m,t} = \varphi \mathrm{mc}_t + \alpha_N p_{m,t} + (1-\alpha_N) p_t,$$

where α_N is given by

$$\alpha_N = \frac{N(\psi-1)(\psi-\varepsilon)}{\psi\left(N^2(\psi-1)-(N-1)\psi\right)+(N-2)\psi\varepsilon+\varepsilon^2}.$$

Turn to the environment where there is nominal rigidity and incomplete information. The problem of a firm that can reset its price becomes

$$\max_{P_{i,m,t}} \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,m,t}\left[P_{i,m,t} Y_{i,m,t+k} - P_{t+k} C(Y_{t+k}) Y_{i,m,t+k}\right],$$

and the linearized first-order condition becomes

$$p^*_{i,m,t} = (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \varphi \mathbb{E}_{i,m,t}[\mathrm{mc}_{t+k}] + (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \mathbb{E}_{i,m,t}[\alpha_N p_{m,t+k} + (1-\alpha_N)p_{t+k}].$$

Under the assumption that all firms share the same information within the market, all newly set prices within a market are identical. Denote the newly set price in market m as $p^*_{m,t}$, and it satisfies

$$p^*_{m,t} = (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \varphi \mathbb{E}_{m,t}[\mathrm{mc}_{t+k}] + (1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k \mathbb{E}_{m,t}[\alpha_N p_{m,t+k} + (1-\alpha_N)p_{t+k}].$$

Denote $\pi_{m,t}$ as the inflation rate in market m. Subtracting $p_{m,t-1}$ from both sides of the equation above leads to

$$\begin{aligned}\pi_{m,t} =& (1-\theta)(1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k\varphi\mathbb{E}_{m,t}[\mathrm{mc}_{t+k}] + \alpha_N(1-\theta)\sum_{k=0}^{\infty}(\chi\theta)^k\mathbb{E}_{m,t}[\pi_{m,t+k}] \\ &+ (1-\theta)(1-\alpha_N)\sum_{k=0}^{\infty}(\chi\theta)^k\mathbb{E}_{m,t}[\pi_{t+k}] + (1-\theta)(1-\alpha_N)(p_{t-1}-p_{m,t-1}).\end{aligned}$$

To proceed, consider the following alternative inflation definition in market m

$$\begin{aligned}\widetilde{\pi}_{m,t} =& (1-\theta)(1-\chi\theta)\sum_{k=0}^{\infty}(\chi\theta)^k\varphi\mathbb{E}_{m,t}[\mathrm{mc}_{t+k}] + \alpha_N(1-\theta)\sum_{k=0}^{\infty}(\chi\theta)^k\mathbb{E}_{m,t}[\widetilde{\pi}_{m,t+k}] \\ &+ (1-\theta)(1-\alpha_N)\sum_{k=0}^{\infty}(\chi\theta)^k\mathbb{E}_{m,t}[\pi_{t+k}].\end{aligned}$$

Since the aggregate inflation under these two models are identical ($\int_m \pi_{m,t} = \int_m \widetilde{\pi}_{m,t}$), we can derive the aggregate inflation dynamics from the latter. By the law of iterated expectations, we have

$$\begin{aligned}\widetilde{\pi}_{m,t} =& \mathbb{E}_{m,t}\left[\frac{(1-\theta)(1-\chi\theta)\varphi\mathrm{mc}_t + (1-\theta)(1-\alpha_N)\pi_t}{1-\chi\theta L^{-1}}\right] + (1-\theta)\alpha_N\mathbb{E}_{m,t}\left[\frac{\widetilde{\pi}_{m,t}}{1-\chi\theta L^{-1}}\right] \\ =& \mathbb{E}_{m,t}\left[\frac{(1-\theta)(1-\chi\theta)\varphi\mathrm{mc}_t + (1-\theta)(1-\alpha_N)\pi_t}{1-\chi\theta L^{-1}}\left(1-\frac{(1-\theta)\alpha_N}{1-\chi\theta L^{-1}}\right)^{-1}\right]\end{aligned}$$

$$= \frac{1}{1-(1-\theta)\alpha_N} \sum_{k=0}^{\infty} \left(\frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \mathbb{E}_{m,t}[(1-\theta)(1-\chi\theta)\varphi \text{mc}_{t+k} + (1-\theta)(1-\alpha_N \pi_{t+k})].$$

Aggregating across markets and using the assumption that firms can observe current inflation, it follows that the aggregate inflation satisfies

$$\pi_t = \kappa \sum_{k=0}^{\infty} \left(\frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \overline{\mathbb{E}}_t[\text{mc}_{t+k}] + \frac{\chi(1-\theta)(1-\alpha_N)}{1-(1-\theta)\alpha_N} \sum_{k=0}^{\infty} \left(\frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \overline{\mathbb{E}}_t[\pi_{t+k+1}],$$

where $\kappa = \frac{(1-\chi\theta)(1-\theta)\varphi}{\theta}$. Mapping the fixed point problem above to our baseline framework, the aggregate outcome is the result of the following forward-looking game

$$a_{i,t} = \varphi \mathbb{E}_{i,t}[\xi_t] + \beta \mathbb{E}_{i,t}[a_{i,t+1}] + \gamma \mathbb{E}_{i,t}[a_{t+1}],$$

where

$$\beta = \frac{\chi\theta}{1-(1-\theta)\alpha_N}, \quad \text{and} \quad \gamma = \frac{\chi(1-\theta)(1-\alpha_N)}{1-(1-\theta)\alpha_N}.$$

with $\beta + \gamma = \chi$. Note that γ is decreasing in α_N. To show that γ is increasing in N, it is sufficient to show that α_N is decreasing in N. When $\psi > \varepsilon > 1$, and $N \geq 2$

$$\begin{aligned}
\frac{\partial \alpha_N}{\partial N} &= \frac{(\psi-1)(\psi-\varepsilon)\left(\psi^2+\varepsilon^2-2\psi\varepsilon-\psi N^2(\psi-1)\right)}{\left(\psi\left(N^2(\psi-1)-(N-1)\psi\right)+(N-2)\psi\varepsilon+\varepsilon^2\right)^2} \\
&< \frac{(\psi-1)(\psi-\varepsilon)\left(\psi^2-\psi+\varepsilon^2-\varepsilon-N^2(\psi^2-\psi)\right)}{\left(\psi\left(N^2(\psi-1)-(N-1)\psi\right)+(N-2)\psi\varepsilon+\varepsilon^2\right)^2} \\
&< 0,
\end{aligned}$$

which completes the proof.

Proof of Proposition 12

The signal process can be represented as

$$\begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} \tau_\varepsilon^{-1/2} & 0 & \frac{1}{1-\rho L} \\ 0 & \tau_u^{-1/2} & \frac{1}{1-\rho L} \end{bmatrix}}_{\equiv \mathbf{M}(L)} \underbrace{\begin{bmatrix} \widehat{\varepsilon}_t \\ \widehat{u}_{i,t} \\ \widehat{\eta}_t \end{bmatrix}}_{\equiv \widehat{\mathbf{s}}_{i,t}}.$$

where $\widehat{\mathbf{s}}_{i,t}$ is a vector of standardized normal random variables. The auto-covariance generating function for the signal process is

$$\mathbf{M}(L)\mathbf{M}'\left(L^{-1}\right) = \frac{1}{(L-\rho)(1-\rho L)} \begin{bmatrix} L + \frac{(L-\rho)(1-\rho L)}{\tau_\varepsilon} & L \\ L & L + \frac{(L-\rho)(1-\rho L)}{\tau_u} \end{bmatrix}.$$

In order to apply the Wiener-Hopf prediction formula we need to obtain the canonical factorization. Let λ be the inside root of the determinant of $\mathbf{M}(L)\mathbf{M}'\left(L^{-1}\right)$

$$\lambda = \frac{1}{2}\left(\frac{\tau_\varepsilon+\tau_u}{\rho}+\frac{1}{\rho}+\rho-\sqrt{\left(\frac{\tau_\varepsilon+\tau_u}{\rho}+\frac{1}{\rho}+\rho\right)^2-4}\right).$$

Then the fundamental representation is given by

$$\mathbf{B}(z)^{-1} = \frac{1}{1-\lambda z}\begin{bmatrix} 1-\frac{\tau_\epsilon\rho+\lambda\tau_u}{\tau_\epsilon+\tau_u}z & \frac{\tau_u(\lambda-\rho)}{\tau_\epsilon+\tau_u}z \\ \frac{\tau_\epsilon(\lambda-\rho)}{\tau_\epsilon+\tau_u}z & 1-\frac{\tau_u\rho+\lambda\tau_\epsilon}{\tau_\epsilon+\tau_u}z \end{bmatrix},$$

$$\mathbf{V}^{-1} = \frac{\tau_\epsilon\tau_u}{\rho(\tau_\epsilon+\tau_u)}\begin{bmatrix} \frac{\tau_u\rho+\lambda\tau_\epsilon}{\tau_u} & \lambda-\rho \\ \lambda-\rho & \frac{\tau_\epsilon\rho+\lambda\tau_u}{\tau_\epsilon} \end{bmatrix},$$

which satisfies

$$\mathbf{B}(L)\mathbf{V}\mathbf{B}'\left(L^{-1}\right) = \mathbf{M}(L)\mathbf{M}'\left(L^{-1}\right).$$

Applying the Wiener-Hopf prediction formula, the forecast of ξ_t is given by

$$\mathbb{E}_{i,t}[\xi_t] = \left[\begin{bmatrix}0 & 0 & \frac{1}{1-\rho L}\end{bmatrix}\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix} = \frac{\lambda\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}}{\rho(1-\lambda L)\left(1-\rho\lambda\right)}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix}.$$

Suppose the policy function is $h_1(L)$ and $h_2(L)$, that is,

$$a_{i,t} = h_1(L)z_t + h_2(L)x_{i,t}.$$

Let $g(L) \equiv h_1(L)+h_2(L)$, and it follows that the aggregate outcome is $a_t = g(L)\xi_t + h_1(L)\epsilon_t$. The forecast about a_{t+1} is given by

$$\mathbb{E}_{i,t}[a_{t+1}] = \left[\begin{bmatrix}\tau_\varepsilon^{-1/2}L^{-1}h_1(L) & 0 & \frac{L^{-1}g(L)}{1-\rho L}\end{bmatrix}\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix}$$

$$= \left\{ \frac{\begin{bmatrix}\left(\left(\rho\tau_u+\lambda\tau_\varepsilon+\lambda\rho\left(\lambda\tau_u+\rho\tau_\varepsilon\right)\right)L-\lambda\rho\left(\tau_u+\tau_\varepsilon\right)\left(1+L^2\right)\right)h_1(L) & \tau_u\left(\lambda-\rho\right)\left(1-\rho\lambda\right)Lh_1(L)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L(L-\lambda)(1-\lambda L)} \right.$$

$$- \frac{\begin{bmatrix}\tau_\varepsilon\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_1(\lambda) & \tau_u\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_1(\lambda)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L(L-\lambda)(1-\lambda L)}$$

$$- \frac{\begin{bmatrix}\rho(L-\lambda)\left(\left(\lambda\tau_u+\rho\tau_\varepsilon\right)L-\left(\tau_u+\tau_\varepsilon\right)\right)h_1(0) & \tau_u\left(\rho-\lambda\right)L\rho(L-\lambda)h_1(0)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L(L-\lambda)(1-\lambda L)}$$

$$\left. + \frac{\lambda\left(\left(1-\rho\lambda\right)g(L)-\left(1-\rho L\right)g(\lambda)\right)\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}}{\rho\left(1-\rho\lambda\right)(L-\lambda)(1-\lambda L)} \right\}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix}.$$

Also, the forecast about $a_{i,t+1} - a_{t+1}$ is

$$\mathbb{E}_{i,t}\left[a_{i,t+1} - a_{t+1}\right] = \left[\begin{bmatrix}0 & \tau_u^{-1/2} L^{-1} h_2(L) & 0\end{bmatrix} \mathbf{M}'\left(L^{-1}\right) \mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+ \mathbf{V}^{-1} \mathbf{B}(L)^{-1} \begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix}$$

$$= \left\{ \frac{\begin{bmatrix}\tau_\varepsilon(\lambda-\rho)(1-\rho\lambda) L h_2(L) & \left(\left(\lambda\tau_u + \rho\tau_\varepsilon + \lambda\rho\left(\rho\tau_u + \lambda\tau_\varepsilon\right)\right) L - \lambda\rho\left(\tau_u + \tau_\varepsilon\right)\left(1 + L^2\right)\right) h_2(L)\end{bmatrix}}{\rho\left(\tau_u + \tau_\varepsilon\right) L (L-\lambda)(1-\lambda L)} \right.$$

$$- \frac{\begin{bmatrix}\tau_\varepsilon(\rho-\lambda)(1-\rho L) L h_2(\lambda) & \tau_u(\rho-\lambda)(1-\rho L) L h_2(\lambda)\end{bmatrix}}{\rho\left(\tau_u + \tau_\varepsilon\right) L (L-\lambda)(1-\lambda L)}$$

$$\left. - \frac{\begin{bmatrix}\tau_\varepsilon(\rho-\lambda) L \rho (L-\lambda) h_2(0) & \rho(L-\lambda)\left(\left(\rho\tau_u + \lambda\tau_\varepsilon\right) L - \left(\tau_u + \tau_\varepsilon\right)\right) h_2(0)\end{bmatrix}}{\rho\left(\tau_u + \tau_\varepsilon\right) L (L-\lambda)(1-\lambda L)} \right\} \begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix}.$$

These two objects are useful for agents to decide their optimal action, which should satisfy the best response function

$$a_{i,t} = \varphi \mathbb{E}_{i,t}[\xi_t] + \beta \mathbb{E}_{i,t}[a_{i,t+1}] + \gamma \mathbb{E}_{i,t}[a_{t+1}] = \varphi \mathbb{E}_{i,t}[\xi_t] + \beta \mathbb{E}_{i,t}[a_{i,t+1} - a_{t+1}] + (\gamma+\beta) \mathbb{E}_{i,t}[a_{t+1}].$$

Substituting the forecast formulas into the best response function, it leads to the following functional equation

$$\mathbf{A}(L) \begin{bmatrix} h_1(L) \\ h_2(L) \end{bmatrix} = \boldsymbol{d}(L),$$

where[49]

$$\mathbf{A}(L) = \begin{bmatrix} 1 - (\gamma+\beta) L^{-1} & -\frac{\gamma\lambda\tau_\varepsilon}{\rho(L-\lambda)(1-\lambda L)} \\ 0 & 1 - \frac{\gamma\lambda\tau_u}{\rho(L-\lambda)(1-\lambda L)} - \beta L^{-1} \end{bmatrix},$$

and

$$\mathbf{D}(L) \equiv \frac{\varphi\lambda\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}'}{\rho(1-\lambda L)(1-\rho\lambda)} - \varphi_1 \frac{(1-\rho L)\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}'}{(L-\lambda)(1-\lambda L)}$$
$$-\varphi_2 \frac{\begin{bmatrix}\left(\lambda\tau_u + \rho\tau_\varepsilon\right) L - \left(\tau_\varepsilon + \tau_u\right) & \tau_u\left(\rho-\lambda\right) L\end{bmatrix}'}{L(1-\lambda L)} - \varphi_3 \frac{\begin{bmatrix}\tau_\varepsilon\left(\rho-\lambda\right) L & \left(\lambda\tau_\varepsilon + \rho\tau_u\right) L - \left(\tau_\varepsilon + \tau_u\right)\end{bmatrix}'}{L(1-\lambda L)},$$

with

$$\varphi_1 = \frac{(\rho-\lambda)\left((\gamma+\beta) h_1(\lambda) + \beta h_2(\lambda)\right)}{\rho\left(\tau_u + \tau_\varepsilon\right)} + (\beta+\gamma)\frac{\lambda g(\lambda)}{\rho(1-\rho\lambda)}, \; \varphi_2 = \frac{\gamma+\beta}{\tau_u + \tau_\varepsilon} h_1(0), \; \varphi_3 = \frac{\beta}{\tau_u + \tau_\varepsilon} h_2(0).$$

[49] We have used the following identities to simplify the expressions

$$\rho\tau_u + \lambda\tau_\varepsilon + \lambda\rho\left(\lambda\tau_u + \rho\tau_\varepsilon\right) + \lambda\tau_\varepsilon(\tau_u + \tau_\varepsilon) = \rho(1+\lambda^2)(\tau_u + \tau_\varepsilon),$$

$$\rho\tau_\varepsilon + \lambda\tau_u + \lambda\rho\left(\lambda\tau_\varepsilon + \rho\tau_u\right) + \lambda\tau_u(\tau_u + \tau_\varepsilon) = \rho(1+\lambda^2)(\tau_u + \tau_\varepsilon).$$

Next note that the determinant of $\mathbf{A}(L)$ is given by

$$\det(\mathbf{A}(L)) = \frac{\lambda\left(-L^3 + \left(\rho + \frac{1}{\rho} + \frac{\tau_u+\tau_\varepsilon}{\rho} + \beta\right)L^2 - \left(1 + \beta\left(\rho + \frac{1}{\rho} + \frac{\tau_u+\tau_\varepsilon}{\rho}\right) + \frac{\gamma\tau_u}{\rho}\right)L + \beta\right)(L - (\gamma+\beta))}{L^2(1-\lambda L)(L-\lambda)},$$

which has four roots ω_1 to ω_4, with $|\omega_4| > 1$ and the others being less than 1 in absolute value. We choose φ_1, φ_2, and φ_3 to remove the inside poles of $h_1(L)$ at ω_1 to ω_3. This leads to the following policy function,

$$h_1(L) = \frac{\varphi}{1-\rho(\beta+\gamma)}\frac{\tau_\varepsilon\vartheta}{\rho(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \quad \text{and} \quad h_2(L) = \frac{\varphi}{1-\rho(\beta+\gamma)}\frac{(1-\rho\vartheta)(\rho-\vartheta)-\vartheta\tau_\varepsilon}{\rho(1-\rho\vartheta)}\frac{1}{1-\vartheta L},$$

where $\vartheta \equiv \omega_4^{-1}$ is the reciprocal of the outside root of the following cubic equation

$$\begin{aligned}
C(z) &= -z^3 + \left(\rho + \frac{1}{\rho} + \frac{\tau_u+\tau_\varepsilon}{\rho} + \beta\right)z^2 - \left(1 + \beta\left(\rho + \frac{1}{\rho} + \frac{\tau_u+\tau_\varepsilon}{\rho}\right) + \frac{\gamma\tau_u}{\rho}\right)z + \beta \\
&= -z^3 + \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + \delta - \gamma\right)z^2 - \left(1 + (\delta-\gamma)\left(\rho + \frac{1}{\rho}\right) + \frac{\delta - \gamma\chi}{\rho\sigma^2}\right)z + \delta - \gamma.
\end{aligned}$$

where the last line using the definition $\sigma^{-2} = \sigma_u^{-2} + \sigma_\epsilon^{-2}$. The aggregate outcome, $a_t = (h_1(L) + h_2(L))\xi_t + h_1(L)\epsilon_t$, is

$$\begin{aligned}
a_t &= \left(1 - \frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\frac{\varphi}{1-\rho(\beta+\gamma)}\xi_t + \frac{\tau_\varepsilon\vartheta}{\rho(1-\rho\vartheta)}\frac{\varphi}{1-\rho(\beta+\gamma)}\frac{1}{1-\vartheta L}\varepsilon_t \\
&\equiv a_t^\xi + v_t.
\end{aligned}$$

In terms of comparative statics, note that

$$\frac{\partial C(\vartheta^{-1})}{\partial\chi} = \frac{\chi}{\rho\sigma^2} > 0.$$

By the same logic in the proof of Proposition 5, it follows that ϑ is decreasing in χ.

Proof of Proposition 13

This follows directly from the analysis in the main text.

Proof of Proposition 14

First, let us prove $g_k < \widehat{g}_k$. Recall that $\{g_k\}$ is given by

$$g_k = \sum_{h=0}^{\infty}\gamma^h\lambda_k\lambda_{k+1}...\lambda_{k+h}\rho_{k+h}.$$

Clearly,

$$0 < g_k < \sum_{h=0}^{\infty}\gamma^h\lambda_k\rho_{k+h} = \widehat{g}_k,$$

which proves the first property. If $\lim_{k\to\infty}\lambda_k = 1$ and $\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}$ exists for all k, then it follows that

$$\lim_{k\to\infty}\frac{\widehat{g}_k}{g_k} = \frac{\lim_{k\to\infty}\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}}{\lim_{k\to\infty}\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}} = 1.$$

Next, let us prove that $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$. By definition,

$$\frac{\widehat{g}_{k+1}}{\widehat{g}_k} = \frac{\lambda_{k+1}}{\lambda_k}\frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}},$$

$$\frac{g_{k+1}}{g_k} = \frac{\lambda_{k+1}}{\lambda_k}\frac{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+2}...\lambda_{k+h+1}\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+1}...\lambda_{k+h}\rho_{k+h}}.$$

Since $\{\lambda_k\}$ is strictly increasing and $\rho_k > 0$, we have

$$\frac{g_{k+1}}{g_k}\Big/\frac{\widehat{g}_{k+1}}{\widehat{g}_k} > \frac{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+1}...\lambda_{k+h}\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+1}...\lambda_{k+h}\rho_{k+h}}\Big/\frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}}.$$

It is sufficient to show that the term on the right-hand side is greater than 1. To proceed, we start with the following observation. If $\theta_1 \ge \theta_2 > 0$, and $\frac{y_2}{y_1+y_2} \ge \frac{x_2}{x_1+x_2}$, then

$$\frac{x_1\theta_1 + x_2\theta_2}{x_1 + x_2} \ge \frac{y_1\theta_1 + y_2\theta_2}{y_1 + y_2}. \tag{82}$$

Note that

$$\frac{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+1}...\lambda_{k+h}\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\lambda_{k+1}...\lambda_{k+h}\rho_{k+h}} = \frac{\rho_{k+1}}{\rho_k}\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+2}}{\rho_k}+\dots},$$

and

$$\frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}} = \frac{\rho_{k+1}}{\rho_k}\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\gamma^2\frac{\rho_{k+2}}{\rho_k}+\dots}.$$

In what follows, we will show by induction that

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+2}}{\rho_k}+\dots} \ge \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\gamma^2\frac{\rho_{k+2}}{\rho_k}+\dots}.$$

We first establish the following inequality

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} \ge \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}}.$$

This inequality is obtained by labeling $\theta_1 = 1, \theta_2 = \frac{\rho_k\rho_{k+2}}{\rho_{k+1}^2}$, $x_1 = y_1 = 1$, $x_2 = \gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}$, and $y_2 = \gamma\frac{\rho_{k+1}}{\rho_k}$, and applying inequality (82). By assumption, $\frac{\rho_k\rho_{k+2}}{\rho_{k+1}^2} \le 1$, which implies $\theta_1 \ge \theta_2 > 0$. Meanwhile,

$$\frac{x_2}{x_1+x_2} = \frac{\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} \le \frac{\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}}{\lambda_{k+1}+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} = \frac{y_2}{y_1+y_2}.$$

Now suppose that

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}} \geq \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}},$$

and we need to show

$$\begin{aligned}&\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}} \\ \geq&\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^{n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\frac{\rho_{k+n}}{\rho_k}}.\end{aligned} \tag{83}$$

Again, to apply (82), let $\theta_1=\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}}$, $\theta_2=\frac{\rho_k\rho_{k+n+1}}{\rho_{k+1}\rho_{k+n}}$, $x_1=1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}$, $x_2=\gamma^n\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}$, $y_1=1+\gamma\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}$, $y_2=\gamma^n\frac{\rho_{k+n}}{\rho_k}$. We have

$$\begin{aligned}&\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}} \\ =&\frac{x_1\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}}+x_2\theta_2}{x_1+x_2} \\ \geq&\frac{x_1\theta_1+x_2\theta_2}{x_1+x_2},\end{aligned}$$

and

$$\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^{n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\frac{\rho_{k+n}}{\rho_k}}=\frac{y_1\theta_1+y_2\theta_2}{y_1+y_2}.$$

To establish (83), it remains to show that $\theta_1\geq\theta_2$ and $\frac{x_2}{x_1+x_2}\leq\frac{y_2}{y_1+y_2}$. Note that

$$\frac{\theta_1}{\theta_2}=\frac{1+\gamma\frac{\rho_{k+1}}{\rho_k}\frac{\rho_{k+2}\rho_k}{\rho_{k+1}^2}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}\frac{\rho_{k+n}\rho_k}{\rho_{k+1}\rho_{k+n-1}}}{\theta_2+\gamma\frac{\rho_{k+1}}{\rho_k}\theta_2+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}\theta_2}.$$

By assumption, $\theta_2<1$ and $\theta_2\leq\frac{\rho_k\rho_{k+i+1}}{\rho_{k+1}\rho_{k+i}}$ when $i\leq n$, which leads to $\theta_1\geq\theta_2$. Also note that

$$\begin{aligned}&\frac{x_2}{x_1+x_2} \\ =&\frac{\gamma^n\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}} \\ \leq&\frac{\gamma^n\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}}{\lambda_{k+1}\ldots\lambda_{k+n}+\gamma\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+1}}{\rho_k}+\ldots+\gamma^{n-1}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^{n}\lambda_{k+1}\ldots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}} \\ =&\frac{y_2}{y_1+y_2}.\end{aligned}$$

This completes the proof that $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$.