Myopia and Anchoring*

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

March 21, 2020

Abstract

We offer a simple toolbox for understanding and quantifying the equilibrium effects of incomplete information. We first develop an observational equivalence between incomplete information and two behavioral distortions: myopia, or extra discounting of the future; and anchoring of the current outcome to the past outcome, as if there were habit. The distortions are larger when GE considerations are more important, reflecting the role of higher-order uncertainty. We next show how to connect the theory to evidence on expectations, giving new guidance on *which* moments of the expectations are best suited for the purpose of disciplining the theory. We finally illustrate the quantitative potential of informational frictions in the context of the New Keynesian Philips and Dynamic IS curves.

*We are grateful to three anonymous referees and an editor 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, Michael Golosov, Jennifer La'O, John Leahy, Kristoffer Nimark, Stephen Morris, Mikkel Plagborg-Moller and seminar participants at the aforementioned conferences, 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 macroeconomics implications of informational frictions and higher-order uncertainty? How do they relate to other sources of "stickiness"? And what is their quantitative potential? In this paper we offer a simple toolbox for addressing these questions.[1]

We establish an observational equivalence between an economy featuring incomplete, heterogenous information and a representative-agent, full-information variant featuring two behavioral distortions: (i) myopia, or extra discounting of future outcomes; and (ii) anchoring of current outcomes to past outcomes, or habit-like behavior. Both of these distortions are larger when GE considerations are more important, as when there are strong input-output linkages behind the New Keynesian Philips curve or a steep Keynesian cross behind the Dynamic IS curve. We extracts lessons for the relation between micro and macro responses; connect to evidence on expectations; and offer a quantitative assessment.

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:

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

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.

In the textbook New Keynesian model, condition (1) could be both the New Keynesian Philips Curve (NKPC), with a_t standing for inflation and ξ_t for the real marginal cost, or 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 familiar, representative-agent benchmarks by letting people have incomplete information about, or an imperfect "understanding" of, the state of Nature. The friction could be the product of dispersed noisy information (Lucas, 1972), or a metaphor for cognitive constraints (Sims, 2003; Tirole, 2015). Crucially, the friction interferes with how agents adjust their beliefs not only about the exogenous impulse but also about the behavior of others (Morris and Shin, 1998; Woodford, 2003).

Main theoretical result. Under appropriate assumptions, the incomplete-information economy is observationally equivalent to a representative-agent one in which condition (1) is modified as follows:

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

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. The one dulls the forward-looking behavior, the other adds a backward-looking element akin to habit or adjustment costs.

[1] Ours is certainly not the first paper to deal with these questions. The key works we build on and our marginal contributions are discussed below. A review of the related literature can be found in Angeletos and Lian (2016).

Both of these distortions increase, not only with the level of noise, but also with a parameter that governs the strategic complementarity or the GE feedback—think of the extent of input-output linkages behind the NKPC, or the slope of the Keynesian cross behind the the Dynamic IS curve. This property offers guidance on how the friction works, how its empirical footprint differs from that of true habit and adjustment costs, and how to interpret and use the available evidence on expectations.

Underlying insights and marginal contribution. Our observational-equivalence result encapsulates two insights that have previously been documented in the literature, albeit in different forms. The first insight is that gradual learning manifests as "stickiness" in the aggregate dynamics (Sims, 2003; Mankiw and Reis, 2002; Woodford, 2003; Nimark, 2008). The second insight is that higher-order uncertainty causes agents to act *as if* they discount forward-looking GE considerations (Angeletos and Lian, 2018).

Relative to this prior literature, our *theoretical* contribution contains: (i) the bypassing of the curse of dimensionality in higher-order beliefs; (ii) the existence, uniqueness and analytical characterization of the equilibrium; and (iii) the observational-equivalence result. This in turn paves the way to our *applied* contribution, which itself consists of: (iv) a few insights about the relation between micro and macro; (v) some guidance on which evidence on expectations is best suited for the purpose of disciplining the theory; and (vi) two quantitative illustrations, one for inflation and another for consumption.

DSGE and micro to macro. Our observational equivalence offers the sharpest to-date illustration of how informational frictions may substitute for the more dubious 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 the Hybrid NKPC.

In addition, by tying both ω_b and ω_f to the GE feedback, our result highlights how these as-if distortions may be endogenous to policies and market structures that regulate this feedback. This has novel implications for, inter alia, how the inertia in inflation may relate to the slope of the Taylor rule and the extent of input-output linkages.

Relatedly, our analysis yields the following, seemingly paradoxical, conclusion: more responsiveness at the micro level comes together with more sluggishness at the macro level. For instance, a larger degree of price flexibility for firms maps to more sluggishness in aggregate inflation, and a higher marginal propensity to consume (MPC) out of income for households maps to more habit-like persistence in aggregate consumption. In both cases, the reason is the larger micro-level responsiveness is associated with a stronger GE feedback and hence a larger bite of higher-order uncertainty.

Connection to evidence on expectations. Our observational equivalence result facilitates a simple quantitative strategy. We show how estimates of ω_f and ω_b can readily be obtained by combining existing knowledge about the relevant GE parameters with an appropriate moment of the average forecasts, such as that estimated in Coibion and Gorodnichenko (2015), or CG for short. This moment is the coefficient of the regression of the average forecast errors on past forecast revisions.

The basic intuition is similar to that articulated in CG: a higher value for the aforementioned coefficient indicates more sluggish adjustment in expectations, or a larger informational friction. But there is a crucial subtlety: 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 masquerade a large friction in terms of the implied values for ω_f and ω_b.

At the same time, we explain why the moment of the average (consensus) forecasts estimated in CG is more "reliable" for our purposes than the individual-level counterpart estimated in Bordalo et al. (2018) and Kohlhas and Broer (2019), or other moments of the individual forecasts such as the cross-sectional dispersion of forecast errors. We further show how our mapping from the CG moment to the distortions of interest may be robust to richer information structures, including endogenous public signals.

Applications to inflation and consumption. We conclude with a few concrete illustrations of how our "toolbox" can be used to assess the macroeconomic effects of informational frictions.

First, we establish, on the basis of the aforementioned mapping from the CG moment to the pair (ω_b, ω_f), that the level of friction implicit in surveys of inflation forecasts is large enough to *alone* rationalize existing estimates of the Hybrid NKPC (Galí and Gertler, 1999; Galí, Gertler, and Lopez-Salido, 2005). This complements prior work that had hypothesized this for *some* level of noise but had not disciplined the theory with expectations data (Nimark, 2008; Woodford, 2003).[2]

Second, we show that, under the lens of the theory, most of the friction's quantitative effect reflects the anchoring of the expectations of the behavior of others (future inflation) rather than the expectations of the fundamental (real marginal cost). This echoes a recurring theme of our paper, which is that the friction works in large part by arresting GE feedbacks.

Third, turning to the demand side, we show that, for a plausible calibration, the level of habit-like persistence induced by the informational friction is comparable to that used in the DSGE literature (Christiano, Eichenbaum, and Evans, 2005; Smets and Wouters, 2007). 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).[3]

Finally, we use a tractable HANK-like extension of our model to shed new light on how heterogeneity may interact with informational frictions. In particular, we show that a positive cross-sectional correlation between MPC and income cyclicality, like that documented empirically in Patterson (2019), amplifies the information-driven sluggishness in the response of aggregate spending to monetary policy.

[2] Related is also the literature on adaptive learning (Sargent, 1993; Evans and Honkapohja, 2012; Marcet and Nicolini, 2003). This literature, too, allows for the anchoring of current outcomes to past outcomes; see, in particular, Carvalho et al. (2017) for an application to inflation. 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.

[3] Complementary in this respect are Carroll et al. (2020) and Auclert, Rognlie, and Straub (2020).

2 Framework

In this section we set up our framework and illustrate its applicability. We also discuss the essence of the friction we are after and the complexity we aim at bypassing.

2.1 Basic ingredients

Time is discrete, indexed by $t \in \{0, 1, ...\}$, and there is a continuum of players, indexed by $i \in [0, 1]$. In each period t, each agent i chooses an action $a_{i,t} \in \mathbb{R}$. The corresponding average action is denoted by a_t. 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 the underlying exogenous fundamental, $\mathbb{E}_{i,t}[\cdot]$ denotes the player's expectation in period t, and (φ, β, γ) are parameters, with $\varphi > 0$, $\gamma \in [0, \delta)$, $\beta = \delta - \gamma$, and $\delta \in (0, 1)$. As it will become clear shortly, $\delta = \beta + \gamma$ parameterizes the agent's overall concern about the future and γ the component of it that reflects GE, or strategic, considerations. We further assume that individual expectations satisfy the law of iterated expectations along with the following limit properties: $\lim_{k\to\infty} \beta^k \mathbb{E}_{i,t}[a_{i,t+k}] = 0$, $\lim_{k\to\infty} \beta^k \mathbb{E}_{i,t}[\xi_{t+k}] = 0$, and $\lim_{k\to\infty} \beta^k \mathbb{E}_{i,t}[a_{t+k}] = 0$.[4] Iterating on condition (3) then yields the following extensive-form representation of the best response of player i in period t:

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

While the recursive form of the best responses seen in condition (3) is relatively more convenient to work with, the extensive form given in condition (4) highlights that a player'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 it also 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 in turn is useful for two purposes. First, it highlights the fixed-point relation between the equilibrium outcome and the expectations of it, whereby higher-order beliefs come into play. And second, it helps nest incomplete-information versions of the building blocks of the New Keynesian model.[5]

[4] The first property can be understood as the transversality condition of a dynamic optimization problem whose Euler condition is given by (3). The second represents a restriction on the fundamental process, trivially satisfied if ξ_t is either bounded or a random walk. The third represents a equilibrium refinement standard in infinite-horizon settings.

[5] 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 resembles the beauty contests considered by Morris and Shin (2002), Woodford (2003), Angeletos and Pavan (2007), Bergemann and Morris (2013), and Huo and Pedroni

2.2 Complete information and beyond

Suppose momentarily that all agents have complete information, meaning that they share the same (although possibly noisy) information and this fact is itself common knowledge. The economy then admits a representative agent. That is, $a_{i,t} = a_t$ and $\mathbb{E}_{i,t} = \mathbb{E}_t$, where $\mathbb{E}_t$ stands for the representative agent's expectation. In this benchmark, condition (3) reduces to the following:

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

which may correspond to the textbook versions of the Dynamic IS and New Keynesian Philips curves, or an elementary asset-pricing equation. Accordingly, 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 2.4 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 amounts to accommodating a friction in how agents reason about the behavior of others, or about GE.

The specific information structure assumed, which aims at maximizing tractability and clarity, is spelled out in Section 3.1. In the remainder of this section, we first illustrate how our setting can nest incomplete-information extensions of the Dynamic IS and the New Keynesian Philips curves; we then discuss the essence of the friction we are after and the complexity we wish to bypass.

2.3 Two Examples: Dynamic IS and NKPC

The (log-linearized) New Keynesian model boils down to two forward-looking equations, the Dynamic IS curve and the New Keynesian Philips Curve (NKPC), along with a specification of monetary policy.

(2019). However, because behavior is *not* forward-looking in these settings, the relevant higher-order beliefs are those regarding the *concurrent* beliefs of others. By contrast, the relevant higher-order beliefs in our setting are those regarding the *future* beliefs of others, as in Allen, Morris, and Shin (2006), Morris and Shin (2006) and Nimark (2017). The best-response structure assumed in the latter set of papers is herein nested by restricting $\beta = 0$. As discussed in Section 2.4, allowing $\beta > 0$ significantly increases the dimensionality and complexity of the relevant higher-order beliefs.

The familiar, 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 subjective discount factor, and $\mathbb{E}_t$ is the expectation of the representative agent. The first equation describes how aggregate spending responds to the real interest rate, the second how inflation responds to the real marginal cost or the output gap. Clearly, both of these conditions are nested in (6), the representative-agent restriction of our model.

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 (8) and that of the firms in (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 $a_t = \pi_t$, $\xi_t = \mathtt{mc}_t$, $\varphi = \kappa$, $\beta = \chi\theta$, $\gamma = \chi(1-\theta)$ 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.[6]

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 using the identity that ties inflation to the average rest price yields condition (9). When all firms share the same, rational expectations, this condition reduces to the familiar, textbook version of the NKPC. Away for that benchmark, condition (9) reveals the precise manner in which expectations of future inflation (the behavior of the 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.

[6] See Section 6 for a perpetual-youth, overlapping-generations extension, which disentangles the MPC from the discount factor and makes clear that γ is given by the former, not the latter. This corroborates the interpretation of γ as a proxy for the slope of the Keynesian cross.

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

An integral part of our contribution is to relate higher-order uncertainty to how agents reason about GE and, subsequently, to elaborate on the empirical content of this perspective. To this goal, we 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.

To illustrate how this decomposition matters when, and only when, information is incomplete, consider the following example. There are two economies, labeled A and B, that share the same process for ξ_t, the same information structure, and 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) becomes $a_t = \varphi \sum_{k=0}^{\infty} \delta^k \overline{\mathbb{E}}_t[\xi_{t+k}]$, 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 $a_t = \varphi \overline{\mathbb{E}}_t\left[\xi_t\right] + \delta \overline{\mathbb{E}}_t\left[a_{t+1}\right]$ and recursive iteration yields

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

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.[7]

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 $t = t_1 < t_2 < ... < t_h = t + k$. It is easy to show that, when $\beta > 0$, the period-t outcome depends on all of these higher-order beliefs.

[7] With complete information (common knowledge), higher-order beliefs collapse to first-order beliefs. With incomplete information (lack of common knowledge), higher-order beliefs depart from first-order beliefs. This formalizes the sense in which incomplete information introduces a friction in how agents reason about others relative to how they reason about themselves.

To further appreciate the added complexity relative to the $\beta = 0$ case, note that, 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, plus $(k-1) \times (k-2) \times (k-3)/6$ types of fourth-order beliefs, and so on. 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 $\beta = 0$.

An integral part of our contribution is the complete bypassing of this complexity. The assumptions that permit this bypassing are spelled out in the next section. 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 The Equivalence Result

This section contains the core of our contribution. We introduce the assumptions that let us bypass the complexity of higher-order beliefs and solve directly the rational-expectations fixed point, proceed to develop our observation-equivalence result, and discuss the main insights encapsulated in it.

3.1 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 player i receives a new private signal in each 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 player's information in period t is the history of signals up to that period.

As anticipated in the previous section, 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.[9] But instead of endogenizing the structure of this noise, we fix it in a way that best serves our purposes.

[8] This abstraction seems the right benchmark for the purposes of our paper, 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.

[9] Indeed, note that the history of the outcome up to, and including, period t is measurable in $\xi^t \equiv (\xi_0, ..., \xi_t)$, which defines the aggregate state of Nature in period t. The corresponding Harsanyi type of agent i is $x_i^t \equiv (x_{i0}, ..., x_{i,t})$. It follows that the assumed signals represent signals not only of the fundamental but also of the outcome.

3.2 Solving the Rational-Expectations Fixed Point

Consider first 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 assumed above, $\mathbb{E}_t[\xi_{t+k}] = \rho^k \xi_t$, for all $t, k \geq 0$. We thus reach the following result, which states that the complete-information outcome follows the same AR(1) process as the fundamental, rescaled by the factor $\frac{\varphi}{1-\rho\delta}$.

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.

To illustrate, consider again the special case in which $\beta = 0$. Recall that this reduces the dimensionality of the relevant higher-order beliefs and gives the outcome as in condition (11). Using the Kalman filter, we can readily show that the first-order belief $\overline{\mathbb{E}}_t[\xi_t]$ follows an AR(2) process:

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

where $\lambda = \rho(1-G)$ and G is the Kalman gain (a function of ρ and σ). 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($h+1, h-1$). In short, beliefs of higher order exhibit increasingly complex dynamics even in the relatively simple case in which $\beta = 0$.

As explained before, the case with $\beta > 0$ is subject to an even greater curse of dimensionality in terms of higher-order beliefs. And yet, this complexity vanishes, as if by magic, once we focus on the rational-expectations fixed point: under our assumptions, the fixed point turns out to be merely an AR(2) process, whose exact form is characterized below.

Proposition 2 (**Equilibrium Characterization**)**.** *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:

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

Condition (16) gives the incomplete-information dynamics as a transformation of the complete-information counterpart. This transformation indexed by the coefficient 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.

In the rest of this section, we first offer some insight into the math behind the result. We then discuss the economics encapsulated in it.

To understand the math, consider the special case in which $\beta = 0$. In this case, the outcome obeys the following law of motion:

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

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 exogenous impulse behind a_t is ξ_t, one can safely guess that a_t inherits the root of ξ_t. Hence, it better be that

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

for *some* b and ϑ. This implies that the three 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. These properties guarantee that the sum in the right-hand side of (18) would be at most an ARMA(3,1) of the 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.

The above step is true for arbitrary b and ϑ. For our guess to be correct, it'd better be that $d = \lambda$ and $c = b$. The first equation guarantees that the MA part and the last AR part cancel out, so that (19) reduces to an AR(2) with the same roots as our initial guess; the second equation makes sure that the scale is also the same. The first equation yields (17); the second yields $b = \left(1 - \frac{\vartheta}{\rho}\right)\left(\frac{\varphi}{1-\rho\delta}\right)$.

This is the crux of how the "magic" of the rational-expectations fixed point works. The proof presented in the Appendix 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, such beliefs follow complicated ARMA processes of ever increasing orders. And yet, the equilibrium continues to follow a simple 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.

In the sequel, we translate these properties in terms of our observational-equivalence result (Proposition 3) and a few complementary results about the interaction of GE effects and incomplete information (Propositions 5 and 6). The following corollary, which proves useful when connecting the theory to evidence on expectations, is also immediate.

Corollary 1. *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 $(\gamma, \delta, \rho, \sigma)$.*

3.3 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 $\omega_f < 1$ and $\omega_b > 0$. The original representative-agent economy is nested with $\omega_f = 1$ and $\omega_b = 0$. Relative to this benchmark, a lower ω_f represents a higher discounting of the future, or less forward-looking behavior; a higher ω_b represents a greater anchoring of the current outcome to the past outcome, or more backward-looking behavior.

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 one provided in Proposition 2. But we prefer the new representation over the earlier one not only because it serves best 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 F.

3.4 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 anticipates a point we make in Section 4.3. 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.

This result also circles backs to our discussion in Section 2.4 regarding the interpretation of higher-order uncertainty as a friction in agents' GE reasoning. We expand on this perspective in Section 4.2 by showing that higher-order uncertainty amounts to slowing down the GE multiplier.

3.5 Robustness

The results presented above depend on plausible but stark assumptions about the process of ξ_t and the information structure. In our view, this limitation is outweighed by the gains in terms of the accomplished theoretical clarity and the simple empirical implementation offered in the next section.

What is more, the key insights regarding myopia, anchoring, and the role of higher-order beliefs are more general. Online Appendix F 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. Such richness is prohibitive in general, but only because agents cannot disentangle information about recent innovations from information about past innovation. 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. Our elegant, observational-equivalence result is lost, but the essence remains.

Two other extensions, better suited for applied purposes, are considered in Appendices B and D. The first one extends the analysis to a multi-variate analogue of condition (4). This extension can handle the three-equation New Keynesian model, the HANK-like variant we consider in Section 6, and a large class of forward-looking, incomplete-information networks. The second extension lets agents observe a public signal on top of the private signals allowed so far. This anticipates an exercise conducted in Section 5, where we show that the accommodation of public information may, paradoxically, intensify the documented frictions once the theory is disciplined by the relevant evidence on expectations.

4 Implications and Connection to Expectation Evidence

In this section, we first discuss various translations of our abstracts results that are useful for applied purposes. We then connect the theory to existing evidence on expectations and develop a simple tool for quantitive purposes, which we make use of in the subsequent sections.

4.1 Connection to DSGE (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, 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 incomplete-information 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 incomplete-information 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. Online Appendix I offers a related result in the context of 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 offer the sharpest available illustration of how informational frictions can substitute for the more ad hoc sources of sluggish adjustment in *all* the equations of standard DSGE models. The basic idea is of course familiar from previous works (e.g., Sims, 2003; Mankiw and Reis, 2002; Woodford, 2003; Nimark, 2008). The added value here is not only the sharpness of the provided representation but also the following lessons.

First, we show that, by intensifying the GE feedback from expectations of future inflation to current inflation, higher price flexibility maps to more *sluggishness* in inflation. Although this prediction may be hard to test, we find it to be an intriguing, new addition to the "paradoxes of flexibility."

Second, 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. A large part of this literature is devoted in boosting the Keynesian multiplier. In the light of our result, the incorporation of incomplete information in that literature may offer a better account of the dynamics of aggregate spending. We illustrate this point in Section 6.

Third, we offer a new rationale for why such distortions may loom large at the macro level even if they are absent at the micro level. Previous work has emphasized that agents may naturally have less information about aggregate shocks than about idiosyncratic shocks (Mackowiak and Wiederholt, 2009). We add that higher-order uncertainty effectively amplifies the bite of the informational friction at the macro level. Together, these insights help merge the pervasive gap between the macroeconomic and microeconomic estimates of habit persistence in consumption, adjustment costs to investment, and backward-looking in inflation. We expand on this point in Appendix E.

Fourth, by tying the macro-level distortions to GE feedbacks, we highlight how the former can be endogenous to policies that regulate the later. For instance, a monetary policy that reacts more aggressively to fluctuations in inflation maps, under the lens of our analysis, to a smaller backward-looking component in the NKPC (that is, to less persistence in inflation for given persistence in the real marginal cost). We discuss evidence supportive of this prediction in Section 5.

Fifth, we build a bridge to a literature that uses bounded rationality to arrest GE multipliers. We expand on this point below.

Last but not least, we offer a simple strategy for quantifying the distortions of interest. We spell out the elements of this strategy in Section 4.3 and put it at work in Section 5.

4.2 Arresting GE Multipliers

Recall that condition (10), which we reproduce below, allowed us to decompose the equilibrium outcome in its PE and GE components:

$$a_t = \underbrace{\varphi \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t [\xi_{t+k}]}_{\text{PE component}} + \underbrace{\gamma \sum_{k=0}^{\infty} \beta^k \overline{\mathbb{E}}_t [a_{t+k+1}]}_{\text{GE component}}. \tag{23}$$

Recasting this decomposition in terms of the Impulse Response Function (IRF) of the outcome to the innovations in the fundamental gives the following identity for all $\tau \geq 0$:

$$\mathtt{IRF}_\tau = \mathtt{PE}_\tau + \mathtt{GE}_\tau$$

where $\mathtt{IRF}_\tau$ is the change in the outcome triggered a one-unit innovation in the fundamental τ periods after its occurrence,[10] and $\mathtt{PE}_\tau$ and $\mathtt{GE}_\tau$ are the corresponding objects for, respectively, the first and the second term seen in condition (23). We can then also define the GE multiplier at lag τ as

$$\mu_\tau \equiv \frac{\mathtt{PE}_\tau + \mathtt{GE}_\tau}{\mathtt{PE}_\tau} = 1 + \frac{\mathtt{GE}_\tau}{\mathtt{PE}_\tau}.$$

Equivalently, $\mathtt{IRF}_\tau = \mu_\tau \cdot \mathtt{PE}_\tau$. With this notation at hand, we can state the following result.

Proposition 6 (**GE multiplier**)**.** *(i) With complete information,*

$$\mu_\tau = \mu^* \equiv 1 + \frac{\rho\gamma}{1-\rho\delta} \quad \forall\tau \tag{24}$$

(ii) With incomplete information,

$$1 < \mu_\tau < \mu^* \ \forall\tau, \quad \mu_\tau \text{ increases with } \tau, \qquad \text{and} \qquad \lim_{\tau\to\infty} \mu_\tau = \mu^*$$

Part (i) provides the complete-information benchmark. For instance, in the context of the Dynamic IS curve, the property that μ^* increases in γ means that the multiplier increases with the marginal propensity to consume. And in the context of the NKPC, it means that the multiplier increases with the degree of price flexibility—a property that has gone largely unnoticed in the literature but plays an important role in our context because it ultimately regulates the bite of the informational friction.[11]

Part (ii) formalizes the sense in which the informational friction arrests, or slows down, the GE feedback. When an innovation occurs, the multiplier takes a relatively small value and the overall effect is close to the PE effect, due to the lack of common knowledge. But as time passes, agents become increasingly convinced that others are also responding, and the multiplier picks up steam. This highlights the dual role played by learning in our setting: learning means, not only the accumulation of information about the exogenous innovation, which maps to a larger PE effect, but also the achievement of higher levels of common knowledge, which maps to a larger GE multiplier.

Let us now contrast the form of GE attenuation accommodated here to those accommodated in Gabaix (2019) and Farhi and Werning (2019). These works amount to setting $\mu_\tau = \mu$ for all τ, for some constant $\mu \in (0, \mu^*)$. That is, they arrest the GE feedback but do not let it pick up force with the passage of time since the shock has occurred. By the same token, they help generate $\omega_f < 1$ but restrict $\omega_b = 0$. These is because these works assume that agents "stubbornly" underestimate the responses of others, whereas our approach lets such underestimation decrease with the passage of time since an innovation has occurred (and the whole process to restart with any new innovation).

[10] For any t and any $\tau \geq 0$, $\mathtt{IRF}_\tau$ equals the objective expectation of a_t conditional on $\eta_{t-\tau} = 1$ and $\eta_s = 0$ for all $s \neq t - \tau$.

[11] For our purposes the key is the dependance of μ^* on γ. But the provided formula for μ^* highlights that the strength of the GE feedback depends, not only on γ, but also on δ and ρ. This is because of the forward-looking nature of the problem.

4.3 Connecting the Theory 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 model to develop a mapping from 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.

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

$$a_{t+k} - \overline{\mathbb{E}}_t[a_{t+k}] = K_{CG}\left(\overline{\mathbb{E}}_t[a_{t+k}] - \overline{\mathbb{E}}_{t-1}[a_{t+k}]\right) + v_{t+k,t} \tag{25}$$

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 be true at the aggregate level, because the past average revision may not be commonly known. To put it more succinctly, $K_{CG} \neq 0$ is possible because the forecast error of one agent can be predictable by the past information of another agent.

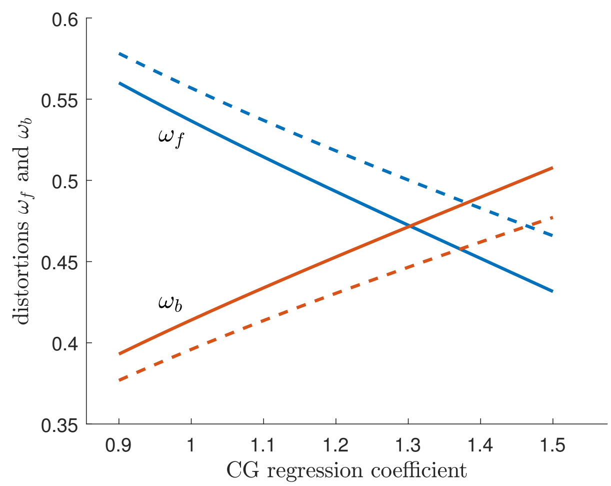
Furthermore, because forecasts adjust sluggishly towards the truth, the theory suggests that K_{CG} oughts to be positive, and the higher the informational friction. To illustrate this, CG treat a_t as an exogenous AR(1) process, assume the same Gaussian signals as we do, and show that in this case $K_{\text{CG}} = \frac{1-g}{g}$, 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.

This logic carries over to our context, where a_t is endogenous and indeed influenced by the informational friction. But there are two complications. First, the identification of σ via K_{CG} is complicated by the fixed point between expectations and outcomes: to invert K_{CG} and find σ, one needs to know how σ feeds into the process of the outcome that agents are trying to forecast. And second, the value of σ (or, equivalently, the associated Kalman gain) is not the "right" measure of the informational friction: insofar γ is large enough, a "tiny" friction as measured in terms of σ could mascaraed a "huge" friction in terms of the deviation of the equilibrium dynamics from its frictionless benchmark.

Our analytical results help handle these complications in an efficient manner. Suppose the analyst has estimates of δ, ρ, and γ from sources other than CG. For instance, in our application to the NKPC (Section 5), δ is the familiar discount factor, ρ is obtained by estimating an AR(1) on a proxy of the real marginal cost; and γ is pinned down by the Calvo parameter, for which there is large empirical literature to draw from. Then, the analyst can use our results to quantify the friction as follows: invert the mapping that relates σ to K_{CG}; plug the result into the mapping that relates σ to (ω_f, ω_b); and obtain a mapping from empirical moment estimated in CG to the macroeconomic distortions of interest.

Figure 1 illustrates how this mapping looks. On the horizontal axis, we vary the value of K_{CG} that may be recovered from running regression (25) on the applicable expectations data. On the vertical axis, we report the predicted values for ω_f and ω_b. Two sets of lines appear in the figure, corresponding to two values for γ.[12] For given γ, a higher K_{CG} maps to more myopia (lower ω_f) and more anchoring (higher ω_b). But a higher γ also maps to larger distortions for given K_{CG}. This echoes our earlier finding that the bite of the information friction increases with the strength of the GE feedback.[13]

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.

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. The same applies for other moments of the average forecasts, such as the persistence of the average forecasts errors estimated in Coibion and Gorodnichenko (2012). But what about moments of the *individual* forecasts?

Consider, in particular, the individual-level counterpart of the CG coefficient, that is, the coefficient of regressing an individual's forecast error on her *own* past revision. As noted earlier, this coefficient is zero under rational expectations: a rational agent's forecast error should be orthogonal to anything measurable in her own past information. Bordalo et al. (2018) and Kohlhas and Broer (2019) instead argue that this coefficient is negative in the data. This suggests the existence of a systematic bias causing forecasts to over-react to new information at the individual level.

[12]The two configurations share the same δ and the same ρ. The specific values used are those described in Section 5.

[13]The difference is that the present result encapsulates the double role that γ plays in the structural interpretation of K_{CG} (the mapping from it to σ) and in the regulation of the relative importance of higher-order uncertainty (the mapping from σ to ω_f and ω_b).

Our own take is that the individual-level evidence is inconclusive: the aforementioned coefficient is negative for the forecasts of some variables (inflation, bond returns) but positive for others (unemployment). In this sense, rational expectations remains valid "on average."

But let us take for granted that that forecasts over-react at the individual level, in the sesen that the aforementioned coefficient is negative. Kohlhas and Broer (2019) attribute such over-reaction to "over-confidence," namely they assume that individuals over-estimate the precision of their information. Bordalo et al. (2018) attribute the same phenomenon to "representativeness," a bias that is a close cousin of over-confidence (at least for the common purposes of those papers and ours).

Motivated by these considerations, in Appendix C we consider an extension of our model that allows agents to be over-confident in the following sense: whereas the actual level of noise remains σ, agents perceive it to be $\hat{\sigma}$, for some $\hat{\sigma} < \sigma$. The opposite case, under-confidence, or $\hat{\sigma} > \sigma$, is also allowed for completeness. The next result summarizes the main lesson for our purposes.

Proposition 7. *Consider an extension in which the perceived level of noise, $\hat{\sigma}$, differs from the actual one, σ. The mapping from K_{CG} to (ω_f, ω_b) is invariant to any potential difference between $\hat{\sigma}$ and σ.*

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 aggregate outcomes in the class of economies we consider. It follows that all our results, including the characterization of (ω_f, ω_b) and K_{CG}, carry over by replacing σ with $\hat{\sigma}$. And by corollary, the mapping from K_{CG} to (ω_f, ω_b) remains the same as before.

The broader lesson is this. The joint properties of the aggregate outcomes and the average forecasts depend only on parameters that govern how agents *perceive* information, whereas the properties of the individual forecasts depend on additional parameters that do not matter for our purposes, such as the actual level of noise. Moments of the average forecasts, such as K_{CG}, therefore serve as "sufficient statistics" for quantifying the distortions of interest.

5 Application to the NKPC

In this section we use the "toolbox" developed in this paper to shed light on the quantitative potential of informational frictions in the context of inflation. In particular, we argue that the theory can not only rationalize existing estimates of the Hybrid NKPC with some level of noise, a point first made in Nimark (2008), but also do so with a level of noise consistent with that inferred from CG's evidence on expectations. We also offer a decomposition in terms of PE and GE channels. We finally illustrate the robustness of our findings to the introduction of public signals.[14]

Operationalizing the theory. Consider the incomplete-information extension of the NKPC presented earlier in Section 2:

$$\pi_t = \kappa \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t [\mathtt{mc}_{t+k}] + \chi(1-\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \overline{\mathbb{E}}_t [\pi_{t+k+1}], \tag{26}$$

where π_t denotes inflation, $\mathtt{mc}_t$ denotes the real marginal cost, $\kappa > 0$ is the slope of the standard NKPC, $\chi \in (0,1)$ is the discount factor, and $\theta \in (0,1)$ is the Calvo parameter. Online Appendix H contains a detailed derivation and a discussion of the underlying assumptions.

Condition (26) allows for a general specification of the process for the real marginal cost and the available signals. But it precludes an analytical characterization, for the reasons explained in Section 2.4. It is also hard to implement empirically in its primitive form, for it requires data on the term structure of the relevant forecasts over long horizons. To make progress, one must either solve for the equilibrium with the help of auxiliary assumptions about the underlying process of the real marginal cost and the available signals, or find some other shortcut.[15]

This is where our paper's toolbox comes handy. As previously noted, our observational-equivalence result allows condition (26) to be reduced to that reported in Corollary 2. By itself, this ties the theory to the existing estimates of the Hybrid NKPC. The analysis of Section 4.3 then further ties the theory to the available evidence on expectations.

To evaluate this tripartite relation between the theory, the existing estimates of the Hybrid NKPC, and the available evidence on expectations, 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, a standard empirical proxy for the real marginal cost and the same as that used in 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

[14] Nimark (2008) foresaw the first part of the application presented below, 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. This aspect distinguishes our application from various other works, including Woodford (2003), Mankiw and Reis (2002), Reis (2006), Kiley (2007), and Matejka (2016). Melosi (2016) estimates an incomplete-information version of the New Keynesian model using a combination of macroeconomic and expectations data, but does not offer the connection to CG and the other lessons offered below; instead, it focuses on a different issue, the signaling role of monetary policy.

[15] Fuhrer (2012) and few other papers have worked with the following equation: $\pi_t = \kappa \mathtt{mc}_t + \chi \overline{\mathbb{E}}_t [\pi_{t+1}]$. Although this is easier to implement empirically, it is invalid under the micro-foundations laid out here. Nimark (2008) and Melosi (2016) share our micro-foundations but use different auxiliary assumptions, which allow for endogenous signals but necessitate numerical methods. They also obtain a different primitive equation than (26), largely because of an oversight that causes expectations of horizons $k \geq 2$ to drop out of this equation; see Online Appendix H for an explanation.

[16] We use seasonally adjusted business sector labor share as proxy for the real marginal cost, from 1947Q1 to 2019Q2. This yields a estimate of ρ equal to 0.9766 or 0.9208 depending on whether we exclude or include a linear trend.

equally under any information structure, it is completely irrelevant for the conclusions drawn below.[17]

Connecting to Existing Estimates of the Hybrid NKPC. The Hybrid NKPC estimated in Galí and Gertler (1999) and Galí, Gertler, and Lopez-Salido (2005) is similar to the one seen in Corollary 2. There is, however, a subtle difference. 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.[18]

A quick, although perhaps not as powerful, test of the theory is therefore whether the existing estimates of the Hybrid NKPC happen to satisfy this restriction. We implement this test in Figure 2. The solid red line depicts the aforementioned restriction: as we increase σ, we travel northwest along this line. The blue 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.[19]

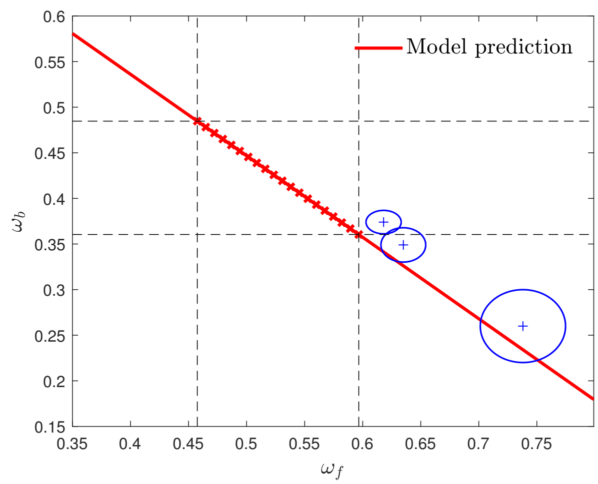


Figure 2: Testing the Theory

Note: The red 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 crossed segment of this line gives the pairs that are consistent with the evidence on expectations provided in Coibion and Gorodnichenko (2015). The three blue crosses represent to the three estimates of the pair (ω_f, ω_b) provided in Table 1 of Galí, Gertler, and Lopez-Salido (2005), and the surrounding disk give the corresponding confidence regions.

As evident in the figure, the theory passes the aforementioned test: the existing estimates of the

[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.

[18] In particular, the restriction imposed by our theory takes the following form: $\omega_f = 1 - \frac{1}{\chi\rho^2}\omega_b$.

[19] 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.

Hybrid NKPC can be rationalized by some level of noise.[20] But is the requisite level of noise empirically plausible? We address this question next.

Bringing in the evidence on expectations. In Section 4.3, we discussed how CG have estimated a key moment of the average forecasts, namely the coefficient K_{CG} of regression (25), and how our results provide a mapping from this moment to the pair (ω_f, ω_b). We now use this mapping to translate the confidence interval of K_{CG} estimated in CG to a segment of the line in Figure 2.

In particular, we take CG's baseline OLS regression, which appears as condition (11) of their paper and concerns the one-year ahead average forecast of inflation in the Survey of Professional Forecasters.[21] As reported in column (1) of Table 1 of that paper, this yields a mean estimate for K_{CG} equal to 1.193, with a standard deviation of 0.185. Translating the 95% confidence interval through the mapping seen in Figure 1 earlier on yields the highlighted segment of the red line in Figure 2 here.

In short, this segment 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. Or, more succinctly, the informational friction implicit in the expectations data may alone account for all the observed inertia in inflation.

Quantitative Bite and 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 black line corresponds to frictionless benchmark, with perfect information. The dashed blue line corresponds to the frictional case, with an informational friction that matches the baseline estimate of K_{CG} reported in CG. The dotted red line is explained later.

As evident by comparing the dashed blue line to the solid black one, 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.

We next use the theory to decompose the overall effect of the informational friction to PE and GE

[20] 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.

[21] The one-year ahead forecast is computed by the average of the nowcast to 3-quarter ahead forecast. Also, while CG rely on the Survey of Professional Forecasters, 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.

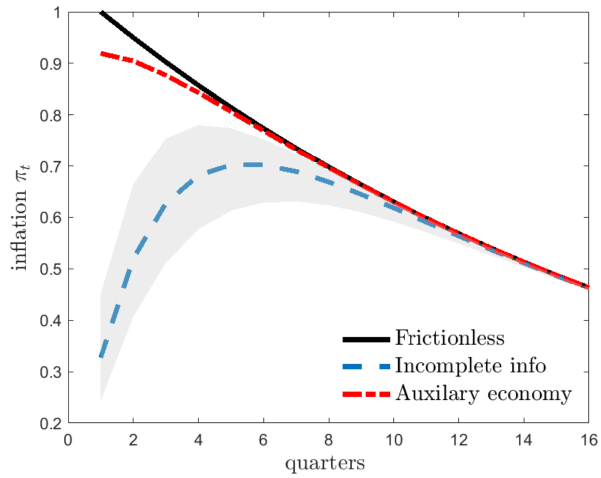


Figure 3: Impulse Response Function of Inflation to Innovations in Real Marginal Cost

components. The dotted red line in Figure 3. represents a counterfactual where the information friction works only through the PE channel. That is, it shuts down the effect of the friction on the expectations of the behavior of others (inflation) and isolates the effect on the the expectations of the fundamental (the real marginal cost). As evident in the figure, this counterfactual delivers a response that is very close the complete-information benchmark and far away from the incomplete-information case. It follows that most of the action in the latter case is in the GE channel, that is, in how the friction anchors the expectations of the behavior of others.[22]

To understand why this is the case, recall that the equilibrium bite of the friction is higher when complete-information GE multiplier is higher to start with. In the present context, this multiplier is quite sizable. In particular, under the textbook calibration employed here, the complete-information value of the GE multiplier is given by

$$\mu^* = 1 + \frac{\rho\chi(1-\theta)}{1-\chi\rho} \approx 7.4,$$

which means that the expectations of future inflation are 6.4 times more important than the expectations of future real marginal costs. This in turn helps explains both why the introduced friction has a quantitatively sizable effect and why most of it works through the GE channel. But it also brings out a broader insight: expectations of future inflation are disproportionally more important than expectations of real marginal costs, or output gaps, in shaping actual inflation dynamics.

Adding a Public Signal. We now explore the robustness of our finding to a different extension, one that preserves rational expectations but adds a public signal. In particular, we first consider the case of an *exogenous* public signal and then turn attention to the case of *endogenous* public signal, namely a

[22] The decomposition offered in Figure 3 mirrors that introduced in Section 2.4. See Online Appendix G for the detailed construction.

noisy statistic of inflation. The first case affords an analytical characterization; the second case requires a numerical approximation but, as shown here, leads to similar conclusions.[23]

Consider the first case, that is, let the firms also observe (and have common knowledge of) a public signal of the form $z_t = \xi_t + \varepsilon_t$, with $\varepsilon_t \sim \mathcal{N}(0, \nu^2)$, in addition the private signals considered so far. Appendix D contains a formal analysis of this case. Here, we briefly discuss the main insights and their quantitative implications.

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 preceding findings above under the requirement that the theory continues to match the available evidence on expectations.

In our benchmark, the CG coefficient was identifying the value of σ, which in turn was pinning down the pair (ω_f, ω_b), or equivalently the equilibrium dynamics. Now, the CG coefficient and the equilibrium dynamics alike depend on two unknown parameters, the precisions $\tau_x \equiv \sigma^{-2}$ and $\tau_z \equiv \nu^{-2}$ 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.[24]

Figure 4 illustrates the implications of these properties for the documented distortions. 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. The upper bound on τ_z turns out to be quite low, simply because evidence in CG points towards high 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.

So far, we have focused the case in which the source of public information is exogenous. We now explore the case where this is endogenous. In particular, we let the public signal be $z_t = \pi_t + \nu_t$, which can be thought of as statistic of inflation contaminated with measurement error.[25] Similar to the exogenous-

[23]We thank an anonymous referee for suggesting this exploration.

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

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

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.

[25]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).

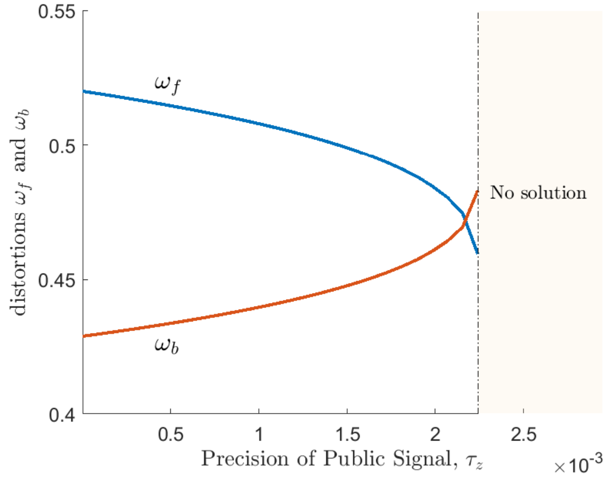


Figure 4: The Role of Public Information

information case, matching the Coibion-Gorodnichenko 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.

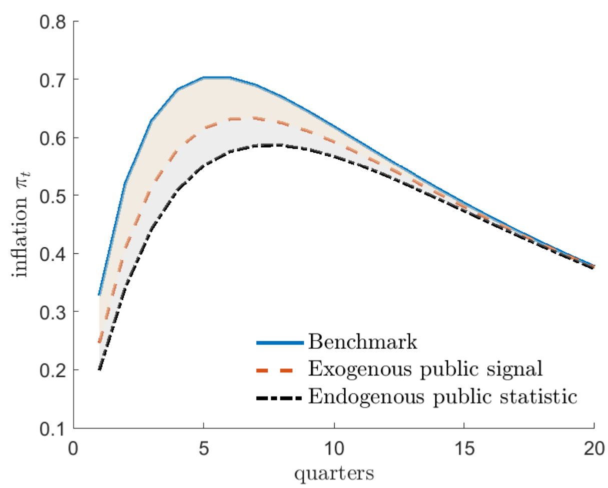


Figure 5: IRF of Inflation, Exogenous vs Endogenous Information

Figure 6 compares the IRF of inflation to real-marginal-cost innovations under three information structures, all required to match the regression coefficient K_{CG} estimated in CG. The blue, solid line corresponds to our benchmark. The red, dashed line correspond to the case in which the public signal is *exogenous* and its precision equals the aforementioned upper bound. The area between this line and the benchmark one spans all the admissible parameterizations of the exogenous-information case. Finally, the black, dotted line corresponds to the case in which the public signal is *endogenous* and its precision equals the appropriate upper bound. The area between this line and the benchmark one spans all the admissible parameterizations of the endogenous-information case.

The main takeaways are two. 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 evidence on expectations in CG.

A third, subtler takeaway is that the endogenous public statistic appears to contribute to more persistence than the exogenous public signal. We find this intriguing and we suspect it is because inflation moves more sluggishly than the fundamental, thus slowing down the information diffusion. 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.

Remarks. As emphasized at the end of Section 4.3, what enters either the CG coefficient or the actual dynamics of inflation according to our theory is the *perceived* level of noise, not the actual one. It follows that the employed method of identifying the friction and the obtained quantitative results are robust to the possibility that agents are over-confident and overreact to private information. For the reasons already explained, this would not have been the case if we had used either the individual-level counterpart of the CG coefficient or the cross-sectional dispersion of forecasts.

One may alternatively try to jointly match multiple moments of the average and individual forecasts at once—or to consider an incomplete-information extension of the entire New Keynesian model and estimate it with maximum likelihood on the basis of data on both outcomes and expectations. How valuable this is depends, in part, on the questions of interest and, in part, on subjective judgements about which approach is more transparent or less susceptible to model mis-specification. We hope that the exercise conducted above offers a simple, useful, first pass. Complementary exercises, which tough focus on different questions, are offered in Nimark (2014) and Melosi (2016).

6 Application to Dynamic IS

Now we turn to the effects of incomplete information on the aggregate demand. As already shown in Corollary 3, through the lens of a perfect-information model, 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).

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 (2017). As in those papers, finite horizons (mortality risk) serve as convenient proxies for liquidity constraints, self-control problems, and other micro-level friction that help explain why most estimates of the MPC in microeconomic data are almost an order of magnitude large 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, of course, we let information be incomplete.

Let us fill in the details. There are two types, or groups, of consumers, indexed by $g \in \{1,2\}$, 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. We allow the households to trade a riskless bond, whose real return is denoted by R_t, and give them access to actuarily fair annuities, so that the effective return to saving is $\frac{R_t}{\varpi_g}$ conditional on survival and zero otherwise. We nevertheless preclude them 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 in a lump sum manner to all consumers, regardless of age. We finally allow the labor income of each group to have a differential exposure to the business cycle.

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 $\beta \in (0,1)$ is the subjective discount factor. Her budget constraint, on the other hand, is given by

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

where $S_{i,g,t;\tau}$ denotes savings in terms of the riskless bond, Y_t denotes aggregate income, 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^* represents the natural rate of output.[26] Using lower-case variables to represent log-deviations from the steady state (e.g., $r_t \equiv \log R_t - \log \chi$), we can express group g's average consumption as follows:

$$c_{g,t} = (1-\chi\varpi_g) \sum_{k=0}^{\infty} (\chi\varpi_g)^k \overline{\mathbb{E}}_t[\phi_g y_{t+k}] - \chi\varpi_g \sum_{k=0}^{\infty} (\chi\varpi_g)^k \overline{\mathbb{E}}_t[r_{t+k}], \quad g \in \{1,2\}, \tag{27}$$

[26]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 and constant monopoly power to get a time-invariant natural rate of output.

where $\overline{\mathbb{E}}_t$ is the average expectation.[27] This is essentially the Permanent Income Hypothesis (see the first term), adapted to allow for variation in the real interest rate (see the second term). This also makes clear that the quantity $\mathrm{m}_g \equiv 1 - \chi\varpi_g$ identifies the MPC of type-g households, and hence also that a higher mortality risk is a metaphor for a higher MPC. We assume that $\phi_1 \geq \phi_2$ and $\varpi_1 \leq \varpi_2$, which means that group 1 has both higher MPC and more cyclical income than group 2.[28]

Aggregate consumption is given by $c_t = \sum_g \pi_g c_{g,t}$. Market clearing imposes $y_t = c_t$. We close the model by treating the real interest rate as an exogenous AR(1) process:

$$r_t = \rho r_{t-1} + \eta_t, \tag{28}$$

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 interpret η_t as a monetary shock.

Equilibrium characterization. In the absence of heterogeneity (i.e., with $\mathrm{m}_1 = \mathrm{m}_2$ and $\phi_1 = \phi_2$), Corollary 3 applies: the dynamics of aggregate consumption obey

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

and a higher level of noise and/or a higher MPC map to a lower ω_f and a higher ω_b.

When instead there is heterogeneity, we proceed as follows. Using market clearing to replace y_t by $y_t = c_t = \sum_g \pi_g c_{g,t}$, and using also the definition of m_g, we can rewrite (27) as follows:

$$c_{g,t} = \mathrm{m}_g \sum_{k=0}^{\infty} (1-\mathrm{m}_g)^k \overline{\mathbb{E}}_t \left[\phi_g \sum \pi_j c_{j,t+k}\right] - (1-\mathrm{m}_g) \sum_{k=0}^{\infty} (1-\mathrm{m}_g)^k \overline{\mathbb{E}}_t[r_{t+k}], \quad g \in \{1,2\}. \tag{29}$$

This allows us to represent the demand block of the economy as incomplete-information network of the kind nested in the analysis of Appendix B. The equilibrium is then solved in the way described in that Appendix.

In the rest of this section, we illustrate the roles of the informational friction and the heterogeneity with a numerical example.[29]

Numerical Example. Figure 6 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 a similar level as that

[27] For simplicity, the precision of information is the same across groups. Therefore, the average expectation within a group is the same as that in the population.

[28] For group incomes to add up to total income, the following restriction must of course hold: $\pi_1\phi_1 + \pi_2\phi_2 = 1$.

[29] When information is complete, the heterogeneity plays no essential role: because of the linearity, the aggregate dynamics are the same as that of a conventional, representative-agent economy. That is, if we denote with c_t^* the complete-information aggregate consumption, we have that c_t^* is proportional to r_t, as in the representative-agent economy.

found in the DSGE literature.[30] 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 $\overline{\mathrm{m}} = .30$, 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; this is the network-like economy described above and solved with the methods developed in Appendix B. In particular, this economy features two groups of consumers, with the first group having both a higher MPC and a more cyclical income: $\mathrm{m}_1 = .55$, $\mathrm{m}_2 = .05$, $\phi_1 = 2$, and $\phi_2 = 0$.

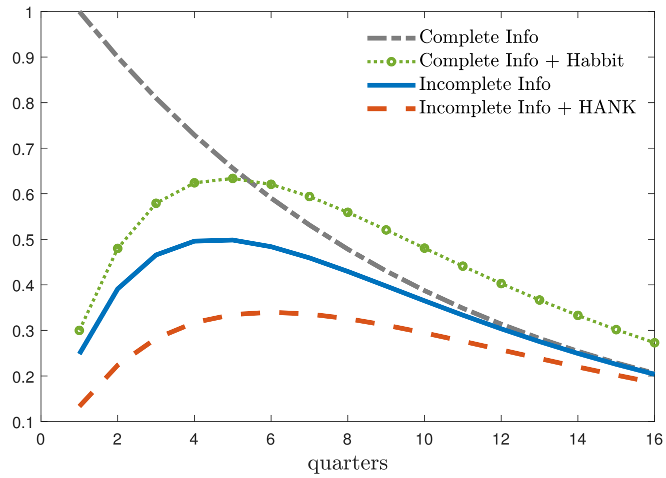


Figure 6: IRF of Consumption, Heterogeneous vs Homogeneous Agents

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 generate 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 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 E, our economy can feature simultaneously a large and front-loaded response to such shocks

[30] In particular, we assume external habit and write the the representative consumer's utility is given by $\log(C_t - b\bar{C}_t)$, 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).

at the micro level, in line with the relevant microeconomic evidence, and at the same time significant 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 have also be dampened and sluggish, contradicting the relevant microeconomic evidence. This idea is pushed further 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, 2016)). 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. We hope that this illustration will invite more work on the intersection of the incomplete-information and HANK literatures.

7 Conclusion

We offered a toolbox for understanding and quantifying the macroeconomic effects of incomplete information, and higher-order uncertainty in particular. This toolbox relied on the use of appropriate, simplifying assumptions. This precluded the analysis of certain issues, such as the endogeneity of the information. But it afforded a high level of theoretical precision, a sharp comparison of incomplete information to bounded rationality, and a straightforward quantitative implementation.

An integral part of our contribution was the connection built to the available evidence on expectations. In particular, we highlighted why the combination of the evidence on the underreaction of the average forecasts provided in Coibion and Gorodnichenko (2012, 2015) and that on the overreaction of individual forecasts provided in Bordalo et al. (2018) and Kohlhas and Broer (2019) speaks in favor of incomplete information over two alternatives, cognitive discounting and level-k thinking, that capture a similar form of GE attention as higher-order uncertainty but abstract from dispersed information. We further clarified why the first kind of evidence may be the most relevant for the quantification of the documented distortions, at least within the class of models under consideration.

The usefulness of the provided tools was illustrated in the context of inflation. For this purpose, we treated the real marginal cost as exogenous and we focused on how incomplete information works through the NKPC alone. Future work may use the multi-variate extension of our results described in Appendix B for more ambitious quantitative endeavors that require the incorporation of the two-way interaction between the NKPC and the Dynamic IS curve.

Another direction for future work may be the extension of our results to more flexible departures from rational expectations. For instance, the analysis can readily be extended to let agents have a misspecified prior not only about the precision of their information but also about the persistence of the fundamental (and hence of the equilibrium outcome, too). When the perceived persistence is lower than the actual one, this extension helps proxy for cognitive discounting as in Gabaix (2019). When the opposite is true, it accommodates a form of over-extrapolative beliefs of a similar flavor as that in Greenwood and Shleifer (2014) and Gennaioli, Ma, and Shleifer (2015). This or other similar extensions could help develop a "unifying" framework, which in turn could be used to address a richer set of empirical regularities and their macroeconomic implications—such as the question of whether average and individual forecasts respond differentially to different shocks and how this matters for macroeconomic outcomes.

References

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

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 and Chen Lian. 2016. "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination." *Handbook of Macroeconomics* 2:1065–1240.

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

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.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris. 2015. "Information and Volatility." *Journal of Economic Theory* 158:427–465.

Bergemann, Dirk and Stephen Morris. 2013. "Robust Predictions in Games with Incomplete Information." *Econometrica* 81 (4):1251–1308.

Bordalo, Pedro, Nicola Gennaioli, Yueran Ma, and Andrei Shleifer. 2018. "Over-reaction in Macroeconomic Expectations." *NBER Working Paper No. 24932* .

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." .

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.

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.

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. 2017. "Monetary Policy, Bounded Rationality, and Incomplete Markets." *NBER Working Paper No. 23281* .

———. 2019. "Monetary Policy, Bounded Rationality, and Incomplete Markets." *American Economic Review* forthcoming.

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

Fuhrer, Jeff. 2012. "The Role of Expectations in Inflation Dynamics." *International Journal of Central Banking* .

Gabaix, Xavier. 2019. "A Behavioral New Keynesian Model." *Harvard mimeo* .

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.

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.

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. 2019. "A Single-Judge Solution to Beauty Contests." *The American Economic Review* forthcoming.

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

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. 2016. "Monetary Policy According to HANK." *NBER Working Paper No. 21897* .

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 Tobias Broer. 2019. "Forecaster (Mis-)Behavior." *IIES miméo* .

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.

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

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.

Marcet, Albert and Juan P Nicolini. 2003. "Recurrent hyperinflations and learning." *American Economic Review* 93 (5):1476–1498.

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* .

Nimark, Kristoffer P. 2014. "Man-Bites-Dog Business Cycles." *American Economic Review* 104 (8):2320–67.

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.

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.

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.

APPENDICES

A Proofs of Propositions in Main Text

This appendix contains the proofs of the results stated in the main text. The proofs of the auxiliary results that appear in the various appendices are collected in Online Appendix K.

Proof of Proposition 1. Follows directly from the analysis in 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

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

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

$$B(L)B(L^{-1}) = \mathbf{M}(L)\mathbf{M}'(L^{-1}) = \frac{\tau_\eta^{-1} + \tau_u^{-1}(1-\rho L)(L-\rho)}{(1-\rho L)(L-\rho)}, \tag{30}$$

which leads to

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

where λ is the inside root of the numerator in equation (30)

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

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:[31]

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

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

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

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

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 $a_{i\,t+1} = h(L)x_{i,t+1}$ 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 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

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

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

$$\zeta_1 = \frac{1}{2\omega_f\delta}\left(1-\sqrt{1-4\delta\omega_f\omega_b}\right) \quad \text{and} \quad \zeta_0 = \frac{\varphi\zeta_1}{\omega_b-\rho\omega_f\delta\zeta_1} \tag{31}$$

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

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

To match the hybrid model, we need

$$\zeta_1 = \vartheta \quad \text{and} \quad \zeta_0 = \left(1-\frac{\vartheta}{\rho}\right)\frac{\varphi}{1-\rho\delta}. \tag{32}$$

Combining (31) and (32), 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:

$$\omega_f = \frac{\delta\rho^2-\vartheta}{\delta(\rho^2-\vartheta^2)} \tag{33}$$

$$\omega_b = \frac{\vartheta(1-\delta\vartheta)\rho^2}{\rho^2-\vartheta^2} \tag{34}$$

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 $\sigma = 0$. The proof of the comparative statics in terms of σ is contained in the proof of Proposition 5.

Proof of Proposition 4 and 5. 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 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 get

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

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. With complete information, the PE component is given by

$$\varphi \sum_{\tau=0}^{\infty} \beta^\tau \mathbb{E}_t [\xi_{t+\tau}] = \frac{\varphi}{1-\rho\beta}\xi_t,$$

and the GE component by

$$\gamma \sum_{\tau=0}^{\infty} \beta^\tau \mathbb{E}_t [a_{t+\tau+1}] = \gamma \sum_{\tau=0}^{\infty} \beta^\tau \mathbb{E}_t \left[\frac{\varphi}{1-\rho\delta}\xi_{t+1}\right] = \frac{\rho\gamma}{1-\rho\delta}\frac{\varphi}{1-\rho\beta}\xi_t,$$

from which part (i) is immediate.

Thus consider part (ii), namely the case with incomplete information. The aggregate outcome is

$$a_t = \frac{\varphi}{1-\rho\delta}\left(1 - \frac{\vartheta}{\rho}\right)\frac{1}{(1-\rho L)(1-\vartheta L)}\eta_t \equiv \frac{\varphi}{1-\rho\delta}\left(1 - \frac{\vartheta}{\rho}\right)\sum_{\tau=0}^{\infty} g_\tau \eta_{t-\tau}$$

The PE component is given by

$$\mathrm{PE}_t = \frac{\varphi}{1-\rho\beta}\left(1 - \frac{\lambda}{\rho}\right)\frac{1}{(1-\rho L)(1-\lambda L)}\eta_t \equiv \frac{\varphi}{1-\rho\beta}\left(1 - \frac{\lambda}{\rho}\right)\sum_{\tau=0}^{\infty} f_\tau \eta_{t-\tau}$$

The sequence of g_τ and f_τ are given by

$$g_\tau = \frac{\rho^{\tau+1} - \vartheta^{\tau+1}}{\rho - \vartheta}, \quad \text{and} \quad f_\tau = \frac{\rho^{\tau+1} - \lambda^{\tau+1}}{\rho - \lambda}.$$

The GE multiplier μ_τ is given by

$$\mu_\tau = \frac{\frac{\varphi}{1-\rho\delta}\left(1 - \frac{\vartheta}{\rho}\right)}{\frac{\varphi}{1-\rho\beta}\left(1 - \frac{\lambda}{\rho}\right)}\frac{g_\tau}{f_\tau} = \left(1 + \frac{\rho\gamma}{1-\rho\delta}\right)\frac{\rho-\vartheta}{\rho-\lambda}\frac{g_\tau}{f_\tau} = \left(1 + \frac{\rho\gamma}{1-\rho\delta}\right)\frac{\rho^{\tau+1} - \vartheta^{\tau+1}}{\rho^{\tau+1} - \lambda^{\tau+1}}.$$

With complete information, $\vartheta = \lambda = 0$, $g_\tau = f_\tau = \rho^\tau$, and we are back to part (i). With incomplete information, we have $\rho > \vartheta > \lambda$. To show μ_τ is increasing in τ, it is equivalent to show that $\frac{g_\tau}{f_\tau}$ is increasing in τ. For $\tau = 0$, $g_0 = f_0 = 1$, and $\frac{g_0}{f_0} = 1$. For $\tau = 1$, $g_1 = \rho + \vartheta$ and $f_1 = \rho + \lambda$, and it follows that

$$\frac{g_1}{f_1} = \frac{\rho+\vartheta}{\rho+\lambda} > \frac{\rho+\lambda}{\rho+\lambda} = 1 = \frac{g_0}{f_0}.$$

For $\tau > 1$, $g_\tau = \rho g_{\tau-1} + \vartheta^\tau$ and $f_\tau = \rho f_{\tau-1} + \lambda^\tau$. To show

$$\frac{g_\tau}{f_\tau} = \frac{\rho g_{\tau-1} + \vartheta^\tau}{\rho f_{\tau-1} + \lambda^\tau} > \frac{g_{\tau-1}}{f_{\tau-1}},$$

it is equivalently to show

$$\vartheta^\tau f_{\tau-1} > \lambda^\tau g_{\tau-1}$$

We prove this by induction. Suppose $\vartheta^\tau f_{\tau-1} > \lambda^\tau g_{\tau-1}$ is true, then

$$\begin{aligned}\vartheta^{\tau+1} f_\tau &= \vartheta^{\tau+1}(\rho f_{\tau-1} + \lambda^\tau) > \vartheta(\rho\lambda^\tau g_{\tau-1} + \lambda^\tau\vartheta^\tau) = \rho\vartheta\lambda^\tau g_{\tau-1} + \lambda^\tau\vartheta^{\tau+1} \\ &> \rho\lambda^{\tau+1} g_{\tau-1} + \lambda^{\tau+1}\vartheta^\tau = \lambda^{\tau+1}(\rho g_{\tau-1} + \vartheta^\tau) = \lambda^{\tau+1} g_\tau,\end{aligned}$$

which completes the induction.

Finally, in the limit as τ goes to ∞,

$$\lim_{\tau\to\infty} \frac{g_\tau}{\lambda_\tau} = \lim_{\tau\to\infty} \frac{\rho-\lambda}{\rho-\vartheta} \frac{\rho^{\tau+1} - \vartheta^{\tau+1}}{\rho^{\tau+1} - \lambda^{\tau+1}} = \frac{\rho-\lambda}{\rho-\vartheta}.$$

Therefore,

$$\lim_{\tau\to\infty} \mu_\tau = \mu^*.$$

Proof of Proposition 7. The proof follows directly from the discussion in Online Appendix C and Proposition 10 therein.

B Multivariate Systems

Corollaries 2 and 3 require that each block of the New Keynesian model—the NKPC and the Dynamic IS curve—is treated in isolation of the other. This is fine for certain purposes, including the exercise conducted in Section 5. As another example, consider the question of how aggregate demand responds to monetary policy. A version of this question can be addressed by assuming that monetary policy induces an AR(1) path for the real interest rate and by studying the mapping from that path to the equilibrium path of aggregate consumption. Corollary 3 then sheds light on how this mapping is distorted by the assumed friction.

For other purposes, however, it is best to allow the "fundamental" in the one block of the model (e.g., the real interest rate faced by the consumers) to depend on the outcome of the other block (e.g., the inflation generated by the joint behavior of the firms). How does this extra layer of GE feedback matters for our main insights and the tractability of our solution?

We address this question by working out a multi-variate extension of our framework. 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:[32]

$$a_{i,g,t} = \varphi_g \mathbb{E}_{i,g,t}[\xi_t] + \beta_g \mathbb{E}_{i,g,t}[a_{i,g,t+1}] + \sum_{j=0}^{n} \gamma_{gj} \mathbb{E}_{i,g,t}[a_{j,t+1}]. \tag{35}$$

The parameters $\{\beta_g\}$ and $\{\gamma_{gj}\}$ help parameterize, once again, PE and GE considerations. The main novelty (and added complexity) is that GE effects now run, not only within groups, but also across groups. This may correspond to the interaction of the two blocks of the New Keynesian model, or more abstractly to a network structure across arbitrary groups of agents. Finally, the parameter φ_g captures the direct exposure of group g to the exogenous shock.[33]

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

$$x_{i,g,t} = \xi_t + u_{i,g,t}, \quad u_{i,g,t} \sim \mathcal{N}(0, \sigma_g^2). \tag{36}$$

where $\sigma_g \geq 0$ parameterizes the noise of group g. Notice that, by allowing σ_g to differ g, we accommodate rich information 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 8. *There exists a unique equilibrium, and the aggregate outcome $a_{g,t}$ of each group g is given*

[32] 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 (35), the results stated below, namely Propositions 8 and 9, continue to hold, modulo a minor adjustment in the cubic that appears in condition (37).

[33] The parameter φ_g is allowed to be zero for some, but not all, g. For instance, if ξ_t represents an aggregate discount-factor shock, it shows up in the Dynamic IS curve but not in NKPC. (And the converse is true of for a markup shock.)

by

$$a_{g,t} = \sum_{j=1}^{n} \psi_{g,j} \left\{ \frac{1-\frac{\vartheta_j}{\rho}}{1-\vartheta_j L} \xi_t \right\} \tag{37}$$

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

$$C(z) = \det\left((\boldsymbol{\delta} - \boldsymbol{\gamma} - \mathbf{I}z)\, \mathbf{diag}\left\{ z^2 - \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma_g^2}\right) z + 1 \right\} - z\, \mathbf{diag}\left\{ \frac{1}{\rho\sigma_g^2} \right\} \boldsymbol{\gamma} \right). \tag{38}$$

Proposition 9. *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:*

$$\boldsymbol{a}_t = \boldsymbol{\varphi}\xi_t + \boldsymbol{\omega}_f \boldsymbol{\delta} \mathbb{E}_t[\boldsymbol{a}_{t+1}] + \boldsymbol{\omega}_b \boldsymbol{a}_{t-1} \tag{39}$$

The proofs of these results are presented at the end of this appendix. One subtlety with the representation offered in condition (39) is that this representation is no more 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 outcome, or the future outcomes of others groups too.[34] 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 (39) is not unique, the equilibrium itself is determinate, and so are its observable properties, which can be directly obtained from Proposition 8.

Proposition 8 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, all the $\{\vartheta_g\}$ have to be jointly determined, and in addition, the outcome of a group depends on all the n different AR(2) processes.

This reflects the rich network structure, or the rich cross-equation feedbacks, allowed by the present extension. To appreciate this, consider the special case in which $\boldsymbol{\beta} = 0$ and $\sigma_g = \sigma$ for all g. In this case, we show in Appendix A that the polynomial given in condition (38) reduces to the product of n quadratics,

[34] 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. Once one of these choices is made, there is no residual indeterminacy.

one for each ϑ_g. Furthermore, the solution for each ϑ_g is in the same manner as in our baseline analysis, 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.

To conclude, the results presented above offer, not only an illustration of the robustness of our main insights to multi-variate systems, but also a straightforward numerical algorithm: one only needs to solve the polynomial seen in condition (38). Contrast this with the complexity of the underlying higher-order beliefs, which now entails not only the issue discussed in Section 2.4 but also the network structure of the higher-order beliefs that each group holds about *other* groups. Proposition 2 may therefore prove useful for future quantitative endeavors.[35]

As mentioned in the main text, the results developed here could assist the study and quantification of incomplete-information versions, not only of the full New Keynesian model, but also of multi-sector models or networks (Bergemann, Heumann, and Morris, 2015; Nimark, Chahrour, and Pitschner, 2019).

C Overconfidence

Consider an extension that lets agents mis-perceive the precision of their information, or equivalently the level (standard deviation) of the noise in their signals: whereas the true level is σ, agents perceive it to be $\hat{\sigma}$, for some $\hat{\sigma} \neq \sigma$. Over-confidence corresponds to $\hat{\sigma} < \sigma$, under-confidence to $\hat{\sigma} > \sigma$.

In a simple context, which abstracts from the fixed-point relation between outcomes and forecasts and amounts to letting $a_t = \xi_t$, Kohlhas and Broer (2019) show that the individual-level counterpart of the CG regression (25) yields a coefficient whose sign is the same as that of the difference between $\hat{\sigma}$ and σ. The basic logic is this. When $\hat{\sigma} = \sigma$, expectations are rational, so individual forecast errors are orthogonal to own past revisions. When instead $\hat{\sigma} < \sigma$, agents over-react to their current information because they over-estimate its precision, so a positive forecast revision today tends to predict a negative forecast error in the future. (And the converse is true if $\hat{\sigma} > \sigma$.)

This logic extends to our context. But the key question of interest to us is how the accommodation of over-confidence influences the fixed-point relation between outcomes and forecasts. The answer is provided by the next result.

Proposition 10. *(i) Propositions 2, 3, 4, 5 and 6 remain valid, modulo the replacement of σ with $\hat{\sigma}$ throughout.*

[35] Like our baseline result, this result depends on abstracting from the endogeneity of information. But unless the analyst is interested *per se* in the effects of a particular endogeneity, this abstraction seems a cost worth paying given the great gain in tractability. Furthermore, the exercise considered in the end of Section 5 indicates our results may offer a useful approximation to settings with endogenous signals as well.

(ii) Any moment of the average forecasts, including the coefficient K_{CG}, is given by the same function as in our baseline model, modulo the replacement of σ with $\hat{\sigma}$.

Apart from establishing the robustness of all our formal results to over-confidence, this result underscores that the documented distortions and their empirical footprint on, say, inflation and average forecasts thereof depends only on the agents' *perception* of the informational friction, not the actual level of noise. The logic is the one articulated in the main text.

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 5). Suppose further that the analyst combines the CG coefficient with the individual-level counterpart estimated in Bordalo et al. (2018) and Kohlhas and Broer (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 if we replace in this argument the CG coefficient with any other moment of the *average* forecasts, such as the persistence of the average forecast errors documented in Coibion and Gorodnichenko (2012), and the moments estimated in Bordalo et al. (2018) and Kohlhas and Broer (2019) with any other moment of the *individual* forecasts, such as the cross-sectional dispersion of the average forecast errors. In this sense, any individual-level evidence drops out of the picture.

Of course, this is not a panacea. The bypassing of the individual-level evidence is justified for our purposes because of a sufficiently narrow focus: we only care about the "macro picture" of how the measured sluggishness in average forecasts maps to myopia and anchoring in aggregate outcomes. But such a bypassing may not be justified if one wishes to have a complete structural model of the joint determination of outcomes and forecasts at both the aggregate and the individual level. In this case, one could use the tools developed here to jointly identify σ and $\hat{\sigma}$ from the combination of all the aforementioned evidence. One could then also ask whether the simple model we have described so far could match a *multitude* of moments of the individual and the average forecasts, or whether further enrichments are needed, such as the distinction between "absolute" and "relative" confidence proposed in Kohlhas and Broer (2019).

We leave these ideas, and a more careful exploration of the expectations data, for future work. In this paper, we instead focus on how the toolbox we have developed here helps quantify the macro-level distortions of interest, and on the importance played by GE feedbacks.

ONLINE APPENDICES

D Public Signals

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

$$z_t = \xi_t + \epsilon_t, \tag{40}$$

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.[36]

Proposition 11. *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.

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

$$C(z) \equiv -z^3 + \left(\rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + (\delta - \gamma)\right) z^2 - \left(1 + (\delta - \gamma)\left(\rho + \frac{1}{\rho}\right) + \frac{\delta - \gamma\chi}{\rho\sigma^2}\right) z + (\delta - \gamma). \tag{41}$$

(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.

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 character-

[36] 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.

ize 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 shown in Section 5, this is not necessarily the case: 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 ($\chi = 0$).

E 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 "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 Mackowiak 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:

$$a_{i,t} = \mathbb{E}_{i,t}[\varphi \xi_{i,t} + \beta a_{i,t+1} + \gamma a_{t+1}] \tag{42}$$

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,

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

It is straightforward to extend the results of Section 3.2 to the current specification. It can thus be shown

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

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 Mackowiak 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).

F 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:

$$\xi_t = \sum_{k=0}^{\infty} \rho_k \eta_{t-k}, \tag{43}$$

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.[38] The second property then permits us, not only to accommodate 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

$$\overline{\mathbb{E}}_t[\xi_t] = \overline{\mathbb{E}}_t\left[\sum_{k=0}^{\infty} \rho_k \eta_{t-k}\right] = \sum_{k=0}^{\infty} f_{1,k}\eta_{t-k}, \qquad \text{with} \qquad f_{1,k} = \lambda_k \rho_k. \tag{44}$$

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 (44) to period $t+1$ and

[38] 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.

taking the period-t average expectation, we get

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

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

$$\overline{\mathbb{F}}_t^h\left[\xi_{t+h-1}\right] = \sum_{k=0}^{\infty} f_{h,k}\eta_{t-k}, \qquad \text{with} \qquad f_{h,k} = \lambda_k\lambda_{k+1}...\lambda_{k+h-1}\rho_{k+h-1}. \tag{45}$$

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 12. *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 Ã 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 Ã 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

$$a_t = \sum_{k=0}^{\infty} g_k \eta_{t-k}, \quad \text{with} \quad g_k = \sum_{h=1}^{\infty} \gamma^{h-1} f_{h,k} = \left\{ \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k \lambda_{k+1} ... \lambda_{k+h-1} \rho_{k+h-1} \right\}. \tag{46}$$

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.[39]

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

[39] 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

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

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 3.3 regarding the role of the persistence of the fundamental.

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

$$a_t = \xi'_t + \gamma' \mathbb{E}_t[a_{t+1}], \tag{47}$$

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.

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 3.4.

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 13. *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:

$$a_t = \varphi \overline{\mathbb{E}}_t[\xi_t] + \gamma \overline{\mathbb{E}}_t[a_{t+1}]. \tag{48}$$

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:

$$g_k = f_{1,k} + \lambda_k \gamma g_{k+1}. \tag{49}$$

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 (47). 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.[40]

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 variant economy. But since the impact is always lower in our economy,[41] 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 4. *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

[40] 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.

[41] Recall, this is by property (i) of Proposition 13.

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 shocks, but only a sum of them, i.e.,

$$x_{i,t,t-k} = \eta_{t-k}^1 + \eta_{t-k}^2 + \epsilon_{i,t,t-k} \quad \forall k.$$

This signal structure is the same as before if agents only care about the sum $\eta_t \equiv \eta_t^1 + \eta_t^2$, 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

$$\overline{\mathbb{E}}_t[\eta_{t-k}^1] = \varpi_1 \lambda_k, \qquad \text{and} \qquad \overline{\mathbb{E}}_t[\eta_{t-k}^2] = \varpi_2 \lambda_k,$$

where the weights ϕ_1 and ϕ_2 depend on the relative volatility of η_t^1 versus η_t^2, 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

$$\overline{\mathbb{E}}_t\left[\phi_1 \xi_t^1\right] = \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k} = \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k}^1 + \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k}^2$$

$$\overline{\mathbb{E}}_t\left[\phi_2 \xi_t^1\right] = \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k} = \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k}^1 + \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k}^2$$

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_t^1. In the frictionless case, $\overline{\mathbb{E}}_t\left[\omega_1 \xi_t^1\right] = \omega_1 \sum_{k=0}^{\infty} \rho_k^1 \eta_{t-k}^1$. The average expectation of ξ_t^1 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_t^2 due to information 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

$$\overline{\mathbb{F}}_t^h\left[\omega_1\xi^1_{t+h-1}+\omega_2\xi^2_{t+h-1}\right]=\sum_{k=0}^{\infty}f_{h,k}\eta_{t-k},\qquad\text{with}\qquad f_{h,k}=\lambda_k\lambda_{k+1}...\lambda_{k+h-1}(\omega_1\phi_1\rho^1_{k+h-1}+\omega_2\phi_2\rho^2_{k+h-1}).$$

Here, we utilize the property that agents cannot separate η_t^1 from η_t^2 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

$$a_t=\sum_{k=0}^{\infty}g_k\eta_{t-k},\quad\text{with}\quad g_k=\sum_{h=1}^{\infty}\gamma^{h-1}f_{h,k}=\left\{\sum_{h=1}^{\infty}\gamma^{h-1}\lambda_k\lambda_{k+1}...\lambda_{k+h-1}(\omega_1\phi_1\rho^1_{k+h-1}+\omega_2\phi_2\rho^2_{k+h-1})\right\}. \tag{50}$$

In contrast, with complete but imperfect information that shares the same first-order belief, the aggregate outcome is

$$a_t=\sum_{k=0}^{\infty}\widehat{g}_k\eta_{t-k},\quad\text{with}\quad\widehat{g}_k=\left\{\sum_{h=1}^{\infty}\gamma^{h-1}\lambda_k(\omega_1\phi_1\rho^1_{k+h-1}+\omega_2\phi_2\rho^2_{k+h-1})\right\}. \tag{51}$$

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 13 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 (45), when agents are rational, the forward higher-order beliefs are

$$\overline{\mathbb{F}}_t^h\left[\xi_{t+h-1}\right]=\sum_{k=0}^{\infty}\lambda_k\lambda_{k+1}...\lambda_{k+h-1}\rho_{k+h-1}\eta_{t-k}.$$

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 $\overline{\mathbb{F}}_t^h\left[\xi_{t+h-1}\right]$ becomes

$$\overline{\mathbb{E}}_t^h\left[\xi_{t+h-1}\right]\equiv\overline{\mathbb{E}}_t\left[\overline{\mathbb{E}}_t[\ldots\overline{\mathbb{E}}_t\left[[\xi_{t+h-1}]\ldots\right]\right]=\sum_{k=0}^{\infty}\lambda_k^h\rho_{t+h-1}\eta_{t-k}.$$

Learning implies $\lambda_{k+1}>\lambda_k$, and the anticipation of learning implies $\lambda_k\lambda_{k+1}...\lambda_{k+h-1}>\lambda_k^h$. 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

$$a_t = \sum_{h=1}^{\infty} \gamma^{h-1} \overline{\mathbb{E}}_t^h [\xi_{t+h-1}],$$

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, $a_{i,t}^0 = \mathbb{E}_{i,t}[\xi_t]$, and the implied aggregate outcome for level-0 thinking is $a_t^0 = \overline{\mathbb{E}}_t[\xi_t]$.

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.[42] 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.

G 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 2.4.

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

[42] 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.

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:

$$\pi_t^* = \varphi\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t\left[\mathtt{mc}_{t+k}|\mathtt{mc}_t\right] + \gamma\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t\left[\pi_{t+k+1}^*|\mathtt{mc}_t\right], \tag{52}$$

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:

$$\tilde{\pi}_t = \varphi\sum_{k=0}^{\infty}\beta^k\overline{\mathbb{E}}_t\left[\mathtt{mc}_{t+k}\right] + \gamma\sum_{k=0}^{\infty}\beta^k\mathbb{E}_t\left[\pi_{t+k+1}^*|\mathtt{mc}_t\right]. \tag{53}$$

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

The dotted red line in Figure 3 corresponds to $\tilde{\pi}_t$. 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.

H 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 5, 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 26 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.

I 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:

$$\text{Cost}_t = \Phi\left(\frac{I_t}{K_{t-1}}\right) \tag{57}$$

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).[43]

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:

$$\text{Cost}_t = \Psi\left(\frac{I_t}{I_{t-1}}\right) \tag{58}$$

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 (57); 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 (58). This illustrates how incomplete information can merge the gap between the different strands of the literature and help reconcile the dominant DSGE practice 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

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

[43] 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 microfoundation of the kind of sluggish investment dynamics featured in the DSGE literature.

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

$$Qq_t + \psi Z z_t = \chi \mathbb{E}_t\left[A\xi_{t+1} + Qq_{t+1} + \psi Z(1+Z)z_{t+1}\right]. \tag{59}$$

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

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

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 (59), 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

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

The following is then immediate.

Proposition 14. *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.

J 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).[44] 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 $\omega_b > 0$:

$$p_t = \mathbb{E}_t[d_{t+1}] + \omega_f \chi \mathbb{E}_t[p_{t+1}] + \omega_b p_{t-1}, \tag{60}$$

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 (60) 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 5, the quantitative

[44] 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.

implications may be similar provided that the theory is disciplined with the relevant evidence on expectations. In the present context, this would mean estimating the analogue of the CG coefficient for the traders' expectations of asset prices and mapping it to an estimate of the coefficients ω_f and ω_b in condition (60). This is another direction for future research.

K Proofs of Auxiliary Results in Appendices

Proof of Proposition 8. 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) $\mathrm{Var}(\overline{\mathbb{E}}_t[X]) \geq \mathrm{Var}(\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_{t+k}[X]])$ for $k \geq 0$; (2) $\mathrm{Var}(aX+bY) \leq (a\sqrt{\mathrm{Var}(X)} + b\sqrt{\mathrm{Var}(Y)})^2$. 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

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

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

$$\begin{aligned}
\mathbb{E}_{i,g,t}[\xi_t] &= \frac{\lambda_g \tau_g}{\rho(1-\rho\lambda_g)} \frac{1}{1-\lambda_g L} x_{i,g,t} \\
\mathbb{E}_{i,g,t}\left[a_{k,t+1}\right] &= \frac{\lambda_g \tau_g}{\rho}\left(\frac{h_k(L)}{(1-\lambda_g L)(L-\lambda_g)} - \frac{h_k(\lambda_g)(1-\rho L)}{(1-\rho\lambda_g)(L-\lambda_g)(1-\lambda_g L)}\right) x_{i,g,t}, \\
\mathbb{E}_{i,g,t}\left[a_{i,g,t+1} - a_{g,t+1}\right] &= \frac{\lambda_g}{\rho}\left(\frac{h_g(L)(L-\rho)}{L(L-\lambda_g)} - \frac{h(\lambda_g)(\lambda_g-\rho)}{\lambda_g(L-\lambda_g)} - \frac{\rho}{\lambda_g}\frac{h_g(0)}{L}\right)\frac{1-\rho L}{1-\lambda_g L} x_{i,g,t}
\end{aligned}$$

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

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

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

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

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 $2n$:

$$\{h_g(0)\}_{g=1}^n,\quad\text{and}\quad\Bigg\{\beta_g h_g(\lambda_g)\Big(\frac{\lambda_g-\rho}{\rho}+\frac{\lambda_g\tau_g}{\rho(1-\rho\lambda_g)}\Big)+\frac{\lambda_g\tau_g}{\rho(1-\rho\lambda_g)}\sum_k\gamma_{g,k}h_k(\lambda_g)\Bigg\}_{g=1}^n. \tag{61}$$

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

$$h_g(L)=\frac{\det\Big[A_1(L)\quad\ldots\quad A_{g-1}(L)\quad d(L)\quad A_g(L)\quad\ldots\quad A_n(L)\Big]}{\det\mathbf{A}(L)}.$$

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 (61), the $2n$ inside roots will be removed. Therefore, the $2n$ constants are solutions to the following system of linear equations:[45]

$$\det\Big[A_1(\zeta_i)\quad\ldots\quad A_{g-1}(\zeta_i)\quad d(\zeta_i)\quad A_g(\zeta_i)\quad\ldots\quad A_n(\zeta_i)\Big]=0,\quad\text{for } i=1,\ldots,n.$$

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

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

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

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

[45] 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.

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

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

Denote $\widetilde{\boldsymbol{a}}_t \equiv \mathbf{Q}\boldsymbol{a}_t$. Because $\mathbf{\Lambda}$ 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 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 9. 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. Denote $\mathbf{Q}$ as $\mathbf{\Lambda}$ as

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

Also denote

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

Note that

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

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

$$\boldsymbol{a}_t = \mathbf{A}\boldsymbol{a}_{t-1} + \mathbf{B}\xi_t \tag{62}$$

In the perfect-information 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 (62) is a solution to the 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

$$\boldsymbol{\omega}_f\boldsymbol{\delta}(\rho\mathbf{B} + \mathbf{A}\mathbf{B}) = \mathbf{B} - \boldsymbol{\varphi}$$
$$\boldsymbol{\omega}_b = \mathbf{A}(\mathbf{I} - \boldsymbol{\omega}_f\boldsymbol{\delta}\mathbf{A})$$

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.

Proof of Proposition 11. 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

$$\mathbb{E}_{i,t}[\xi_t] = \left[\begin{bmatrix} 0 & 0 & \frac{1}{1-\rho L}\end{bmatrix}\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+ \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix} = \frac{\lambda\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}}{\rho(1-\lambda L)(1-\rho\lambda)}\begin{bmatrix} z_t \\ x_{i,t}\end{bmatrix}.$$

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

$$\mathbb{E}_{i,t}[a_{t+1}] = \left[\begin{bmatrix}\tau_\varepsilon^{-1/2}L^{-1}h_1(L) & 0 & \frac{L^{-1}g(L)}{1-\rho L}\end{bmatrix}\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\begin{bmatrix}z_t\\x_{i,t}\end{bmatrix}$$

$$= \left\{\frac{\begin{bmatrix}\left(\left(\rho\tau_u+\lambda\tau_\varepsilon+\lambda\rho\left(\lambda\tau_u+\rho\tau_\varepsilon\right)\right)L-\lambda\rho\left(\tau_u+\tau_\varepsilon\right)\left(1+L^2\right)\right)h_1(L) & \tau_u\left(\lambda-\rho\right)\left(1-\rho\lambda\right)Lh_1(L)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}\right.$$

$$-\frac{\begin{bmatrix}\tau_\varepsilon\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_1(\lambda) & \tau_u\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_1(\lambda)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}$$

$$-\frac{\begin{bmatrix}\rho\left(L-\lambda\right)\left(\left(\lambda\tau_u+\rho\tau_\varepsilon\right)L-\left(\tau_u+\tau_\varepsilon\right)\right)h_1(0) & \tau_u\left(\rho-\lambda\right)L\rho\left(L-\lambda\right)h_1(0)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}$$

$$\left.+\frac{\lambda\left(\left(1-\rho\lambda\right)g(L)-\left(1-\rho L\right)g(\lambda)\right)\begin{bmatrix}\tau_\varepsilon & \tau_u\end{bmatrix}}{\rho\left(1-\rho\lambda\right)\left(L-\lambda\right)\left(1-\lambda L\right)}\right\}\begin{bmatrix}z_t\\x_{i,t}\end{bmatrix}.$$

Also, the forecast about $a_{i,t+1} - a_{t+1}$ is

$$\mathbb{E}_{i,t}\left[a_{i,t+1}-a_{t+1}\right] = \left[\begin{bmatrix}0 & \tau_u^{-1/2}L^{-1}h_2(L) & 0\end{bmatrix}\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'\left(L^{-1}\right)^{-1}\right]_+\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\begin{bmatrix}z_t\\x_{i,t}\end{bmatrix}$$

$$= \left\{\frac{\begin{bmatrix}\tau_\varepsilon\left(\lambda-\rho\right)\left(1-\rho\lambda\right)Lh_2(L) & \left(\left(\lambda\tau_u+\rho\tau_\varepsilon+\lambda\rho\left(\rho\tau_u+\lambda\tau_\varepsilon\right)\right)L-\lambda\rho\left(\tau_u+\tau_\varepsilon\right)\left(1+L^2\right)\right)h_2(L)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}\right.$$

$$-\frac{\begin{bmatrix}\tau_\varepsilon\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_2(\lambda) & \tau_u\left(\rho-\lambda\right)\left(1-\rho L\right)Lh_2(\lambda)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}$$

$$\left.-\frac{\begin{bmatrix}\tau_\varepsilon\left(\rho-\lambda\right)L\rho\left(L-\lambda\right)h_2(0) & \rho\left(L-\lambda\right)\left(\left(\rho\tau_u+\lambda\tau_\varepsilon\right)L-\left(\tau_u+\tau_\varepsilon\right)\right)h_2(0)\end{bmatrix}}{\rho\left(\tau_u+\tau_\varepsilon\right)L\left(L-\lambda\right)\left(1-\lambda L\right)}\right\}\begin{bmatrix}z_t\\x_{i,t}\end{bmatrix}.$$

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[46]

$$\mathbf{A}(L) = \begin{bmatrix} 1-(\gamma+\beta)L^{-1} & -\frac{\gamma\lambda\tau_\varepsilon}{\rho(L-\lambda)(1-\lambda L)} \\ 0 & 1-\frac{\gamma\lambda\tau_u}{\rho(L-\lambda)(1-\lambda L)} - \beta L^{-1} \end{bmatrix},$$

and

$$\mathbf{D}(L) \equiv \frac{\varphi\lambda\left[\tau_\varepsilon \quad \tau_u\right]'}{\rho(1-\lambda L)(1-\rho\lambda)} - \varphi_1 \frac{(1-\rho L)\left[\tau_\varepsilon \quad \tau_u\right]'}{(L-\lambda)(1-\lambda L)}$$
$$-\varphi_2 \frac{\left[(\lambda\tau_u+\rho\tau_\varepsilon)L-(\tau_\varepsilon+\tau_u) \quad \tau_u(\rho-\lambda)L\right]'}{L(1-\lambda L)} - \varphi_3 \frac{\left[\tau_\varepsilon(\rho-\lambda)L \quad (\lambda\tau_\varepsilon+\rho\tau_u)L-(\tau_\varepsilon+\tau_u)\right]'}{L(1-\lambda L)},$$

with

$$\varphi_1 = \frac{(\rho-\lambda)((\gamma+\beta)h_1(\lambda)+\beta h_2(\lambda))}{\rho(\tau_u+\tau_\varepsilon)} + (\beta+\gamma)\frac{\lambda g(\lambda)}{\rho(1-\rho\lambda)},\ \varphi_2 = \frac{\gamma+\beta}{\tau_u+\tau_\varepsilon}h_1(0),\ \varphi_3 = \frac{\beta}{\tau_u+\tau_\varepsilon}h_2(0).$$

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

$$\begin{aligned} a_t &= \left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\frac{\varphi}{1-\rho(\beta+\gamma)}\xi_t+\frac{\tau_\varepsilon\vartheta}{\rho(1-\rho\vartheta)}\frac{\varphi}{1-\rho(\beta+\gamma)}\frac{1}{1-\vartheta L}\varepsilon_t \\ &\equiv a_t^\xi+v_t \end{aligned}$$

[46]We have used the following identities to simplify the expressions

$$\rho\tau_u+\lambda\tau_\varepsilon+\lambda\rho\left(\lambda\tau_u+\rho\tau_\varepsilon\right)+\lambda\tau_\varepsilon(\tau_u+\tau_\varepsilon)=\rho(1+\lambda^2)(\tau_u+\tau_\varepsilon),$$
$$\rho\tau_\varepsilon+\lambda\tau_u+\lambda\rho\left(\lambda\tau_\varepsilon+\rho\tau_u\right)+\lambda\tau_u(\tau_u+\tau_\varepsilon)=\rho(1+\lambda^2)(\tau_u+\tau_\varepsilon).$$

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 12. This follows directly from the analysis in the main text.

Proof of Proposition 13. 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 remains 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

$$\frac{x_1\theta_1 + x_2\theta_2}{x_1 + x_2} \ge \frac{y_1\theta_1 + y_2\theta_2}{y_1 + y_2}$$

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

$$\frac{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h+1}}{\sum_{h=0}^{\infty}\gamma^h\rho_{k+h}} = \frac{\rho_{k+1}}{\rho_k}\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\gamma^2\frac{\rho_{k+2}}{\rho_k}+\dots}$$

Based on the observation, we will show that

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\gamma^2\lambda_{k+1}\lambda_{k+2}\frac{\rho_{k+2}}{\rho_k}+\dots} \geq \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\gamma^2\frac{\rho_{k+3}}{\rho_{k+1}}+\dots}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\gamma^2\frac{\rho_{k+2}}{\rho_k}+\dots}$$

by induction. We first establish the following

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} \geq \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}}$$

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 $y_2 = \gamma\frac{\rho_{k+1}}{\rho_k}$. By assumption, $\frac{\rho_k\rho_{k+2}}{\rho_{k+1}^2} \leq 1$. Meanwhile,

$$\frac{x_2}{x_1+x_2} = \frac{\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} \leq \frac{\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}}{\lambda_{k+1}+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}} = \frac{y_2}{y_1+y_2}$$

Now suppose that

$$\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}} \geq \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}}$$

We want to show

$$\begin{aligned}&\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^n\lambda_{k+1}\dots\lambda_{k+n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^n\lambda_{k+1}\dots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}}\\ \geq&\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^n\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^n\frac{\rho_{k+n}}{\rho_k}}\end{aligned}$$

Let $\theta_1 = \frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}}$, $\theta_2 = \frac{\rho_k\rho_{k+n+1}}{\rho_{k+1}\rho_{k+n}}$, $x_1 = 1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}$, $x_2 = \gamma^n\lambda_{k+1}\dots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}$, $y_1 = 1+\gamma\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_k}$, $y_2 = \gamma^n\frac{\rho_{k+n}}{\rho_k}$. We have

$$\begin{aligned}&\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^n\lambda_{k+1}\dots\lambda_{k+n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}+\gamma^n\lambda_{k+1}\dots\lambda_{k+n}\frac{\rho_{k+n}}{\rho_k}}\\ =&\frac{x_1\frac{1+\gamma\lambda_{k+1}\frac{\rho_{k+2}}{\rho_{k+1}}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n}}{\rho_{k+1}}}{1+\gamma\lambda_{k+1}\frac{\rho_{k+1}}{\rho_k}+\dots+\gamma^{n-1}\lambda_{k+1}\dots\lambda_{k+n-1}\frac{\rho_{k+n-1}}{\rho_k}}+x_2\theta_2}{x_1+x_2}\\ \geq&\frac{x_1\theta_1+x_2\theta_2}{x_1+x_2}\end{aligned}$$

and

$$\frac{1+\gamma\frac{\rho_{k+2}}{\rho_{k+1}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n}}{\rho_{k+1}}+\gamma^{n}\frac{\rho_{k+n+1}}{\rho_{k+1}}}{1+\gamma\frac{\rho_{k+1}}{\rho_{k}}+\ldots+\gamma^{n-1}\frac{\rho_{k+n-1}}{\rho_{k}}+\gamma^{n}\frac{\rho_{k+n}}{\rho_{k}}}=\frac{y_1\theta_1+y_2\theta_2}{y_1+y_2}$$

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 $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$.