

# Online Appendix for Myopia and Anchoring

By GEORGE-MARIOS ANGELETOS AND ZHEN HUO\*

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

## A. PROOFS OF PROPOSITIONS IN MAIN TEXT

### *Proof of Proposition 1*

The proof follows from the main text.

### *Proof of Proposition 2*

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

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

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

$$(A1) \quad B(L)B(L^{-1}) = \mathbf{M}(L)\mathbf{M}'(L^{-1}) = \frac{\tau_\eta^{-1} + \tau_u^{-1}(1 - \rho L)(L - \rho)}{(1 - \rho L)(L - \rho)}.$$

This condition implies that

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

where  $\lambda$  is the inside root of the numerator in the last term of equation (A1)

$$(A2) \quad \lambda = \frac{1}{2} \left[ \rho + \frac{1}{\rho} \left( 1 + \frac{\tau_u}{\tau_\eta} \right) - \sqrt{\left( \rho + \frac{1}{\rho} \left( 1 + \frac{\tau_u}{\tau_\eta} \right) \right)^2 - 4} \right].$$

The forecast of a random variable

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

\* Angeletos: Department of Economics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139 (email: angelet@mit.edu); Huo: Department of Economics, Yale University, 28 Hillhouse Ave, New Haven, CT 06511 (email: zhen.huo@yale.edu).

can be obtained by using the Wiener-Hopf prediction formula:<sup>1</sup>

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

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

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

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

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

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

Consider the forecast of the fundamental. Note that

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

from which it follows that

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

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

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

and the forecasts are

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

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

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  $[G_1(L) + (\beta + \gamma)G_2(L) + \beta G_3(L)]x_{i,t}$ . 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

$$h(z) = \varphi \frac{\lambda \tau_u}{\rho \tau_\eta} \frac{1}{1 - \rho\lambda} \frac{1}{1 - \lambda z} + (\beta + \gamma) \frac{\lambda \tau_u}{\rho \tau_\eta} \left( \frac{h(z)}{(1 - \lambda z)(z - \lambda)} - \frac{h(\lambda)(1 - \rho z)}{(1 - \rho\lambda)(z - \lambda)(1 - \lambda z)} \right)$$

<sup>1</sup>See [Whittle \(1963\)](#) for more details about Wiener-Hopf prediction formula.

$$+ \beta \frac{\lambda}{\rho} \left( \frac{h(z)(z - \rho)}{z(z - \lambda)} - \frac{h(\lambda)(\lambda - \rho)}{\lambda(z - \lambda)} - \frac{\rho h(0)}{\lambda z} \right) \frac{1 - \rho z}{1 - \lambda z},$$

which can be transformed as

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

where

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

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  $\vartheta^{-1}$  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  $\vartheta^{-1}$  is larger than 1.

Finally, to show that  $\vartheta$  is less than  $\rho$ , 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  $\vartheta^{-1}$  is larger than  $\rho^{-1}$ , 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

$$(A3) \quad \zeta_1 = \frac{1}{2\omega_f\delta} \left(1 - \sqrt{1 - 4\delta\omega_f\omega_b}\right) \quad \text{and} \quad \zeta_0 = \frac{\varphi\zeta_1}{\omega_b - \rho\omega_f\delta\zeta_1},$$

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

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

To match the hybrid model, we need

$$(A4) \quad \zeta_1 = \vartheta \quad \text{and} \quad \zeta_0 = \left(1 - \frac{\vartheta}{\rho}\right) \frac{\varphi}{1 - \rho\delta}.$$

Combining (A3) and (A4), and solving for the coefficients of  $\omega_f$  and  $\omega_b$ , we infer that the two economies generate the same dynamics if and only if the following two conditions hold:

$$(A5) \quad \omega_f = \frac{\delta\rho^2 - \vartheta}{\delta(\rho^2 - \vartheta^2)},$$

$$(A6) \quad \omega_b = \frac{\vartheta(1 - \delta\vartheta)\rho^2}{\rho^2 - \vartheta^2}.$$

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

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

#### *Proof of Propositions 4 and 5*

To prove the comparative statics, we first show that  $\omega_f$  is decreasing in  $\vartheta$  and  $\omega_b$  is increasing in  $\vartheta$ . This can be verified as follows

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

Now to prove Proposition 5, it is sufficient to show that  $\vartheta$  is increasing in  $\gamma$ . Note that

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

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,  $\vartheta^{-1}$  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  $\vartheta$ .

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  $\gamma$ , and evaluating that derivative at  $z = \vartheta^{-1}$ , we obtain

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

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

Similarly, taking derivative with respect to  $\tau_u$ , 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  $\vartheta$  is also increasing in  $\sigma$ .

### *Proof of Proposition 6*

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

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

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

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

and it follows that

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

The moment  $K_{CG}$  can be computed as

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

which is the formula given in the Proposition.

Consider next the partial derivatives of  $K_{CG}$  with respect to  $\lambda$  and  $\vartheta$ :

$$(A7) \quad \frac{\partial K_{CG}}{\partial \lambda} = \frac{\left( \begin{array}{l} \theta^4 \lambda^2 \rho (\lambda^2 (\rho^2 + 1) - 4\lambda\rho + \rho^2 + 1) - \theta^3 (4\lambda^3 \rho^3 + \lambda^2 (1 - 6\rho^2) + \rho^2), \\ + \theta^2 \rho (\lambda^2 (6\rho^2 - 1) - 4\lambda\rho - \rho^2 + 1) + 2\theta\rho^2(1 - 2\lambda\rho) + \rho^3 \end{array} \right)}{(1 - \theta\lambda)^2(\rho - \lambda)^2(\theta + \rho - \theta\lambda\rho)^2}.$$

$$(A8) \quad \frac{\partial K_{CG}}{\partial \vartheta} = -\frac{\theta\lambda(2\rho(1 - \theta\lambda) + \theta)}{(1 - \theta\lambda)^2(\theta + \rho - \theta\lambda\rho)^2}$$

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

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

Because  $\vartheta$  increases in  $\gamma$  and  $\lambda$  is invariant in  $\gamma$ , we immediately have that  $K_{CG}$  is decreasing in  $\gamma$ , as stated in the Proposition.

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

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

$$(A9) \quad \frac{dK_{CG}}{d\lambda} = \frac{\partial K_{CG}}{\partial \lambda} + \frac{\partial K_{CG}}{\partial \vartheta} \frac{\partial \vartheta}{\partial \lambda},$$

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

$$(A10) \quad \rho(1 - \beta\theta)(\theta - \lambda)(1 - \theta\lambda) - \gamma\theta^2(\rho - \lambda)(1 - \lambda\rho) = 0.$$

We then apply the Implicit Function Theorem to obtain

$$(A11) \quad \frac{\partial \vartheta}{\partial \lambda} = \frac{\rho(\beta\theta - 1)(\theta^2 - 2\theta\lambda + 1) + \gamma\theta^2(-2\lambda\rho + \rho^2 + 1)}{\rho(\beta(3\theta^2\lambda - 2\theta(\lambda^2 + 1) + \lambda) - 2\theta\lambda + \lambda^2 + 1) - 2\gamma\theta(\lambda^2\rho - \lambda(\rho^2 + 1) + \rho)}.$$

Next, we solve (A10) for  $\gamma$ :

$$(A12) \quad \gamma = \Gamma(\theta; \lambda, \beta, \rho) \equiv \frac{\rho(1 - \beta\theta)(\theta - \lambda)(1 - \theta\lambda)}{\theta^2(\rho - \lambda)(1 - \lambda\rho)}.$$

This identifies the value of  $\gamma$  that induces as an equilibrium any given value form  $\vartheta$  in the admissible range  $[\lambda, \rho]$ . Replacing this value for  $\gamma$  into (A11) allows us to re-express the latter as follows:

$$(A13) \quad \frac{\partial \vartheta}{\partial \lambda} = \frac{\theta(\lambda^2 - 1)(\beta\theta - 1)(\theta^2\rho - \theta(\rho^2 + 1) + \rho)}{(\lambda - \rho)(\lambda\rho - 1)(\beta\theta^3\lambda - \theta(\beta\lambda + \lambda^2 + 1) + 2\lambda)}.$$

Combining the above with (A7), (A8), and (A9), we obtain the following result:

$$(A14) \quad \frac{dK_{CG}}{d\lambda} = \frac{\left( \begin{array}{l} \beta\lambda^5\rho^4\theta^6 + \beta\lambda^3\rho^4\theta^6 - 5\beta\lambda^4\rho^3\theta^6 - \beta\lambda^2\rho^3\theta^6 + \beta\lambda^5\rho^2\theta^6 + 5\beta\lambda^3\rho^2\theta^6 - \beta\lambda^4\rho\theta^6 - \beta\lambda^2\rho\theta^6 \\ -3\beta\lambda^4\rho^4\theta^5 + \beta\lambda^2\rho^4\theta^5 + 3\beta\lambda^3\rho^3\theta^5 + \beta\lambda^2\theta^5 + 3\beta\lambda^4\rho^2\theta^5 - 2\beta\lambda^2\rho^2\theta^5 - 3\beta\lambda^3\rho\theta^5 - \lambda^6\rho^4\theta^4 \\ -\beta\lambda^5\rho^4\theta^4 - 2\lambda^4\rho^4\theta^4 + 4\beta\lambda^3\rho^4\theta^4 - \lambda^2\rho^4\theta^4 - 2\beta\lambda\rho^4\theta^4 + \beta\lambda^3\theta^4 + 5\lambda^5\rho^3\theta^4 + 3\beta\lambda^4\rho^3\theta^4 + 8\lambda^3\rho^3\theta^4 \\ -3\beta\lambda^2\rho^3\theta^4 - \lambda\rho^3\theta^4 - \lambda^6\rho^2\theta^4 - \beta\lambda^5\rho^2\theta^4 - 8\lambda^4\rho^2\theta^4 - 2\beta\lambda^3\rho^2\theta^4 - 4\lambda^2\rho^2\theta^4 + 2\beta\lambda\rho^2\theta^4 + \rho^2\theta^4 \\ +\lambda^5\rho\theta^4 - \beta\lambda^4\rho\theta^4 + 3\lambda^3\rho\theta^4 + 5\lambda^5\rho^4\theta^3 + 3\beta\lambda^4\rho^4\theta^3 + 2\lambda^3\rho^4\theta^3 - 2\beta\lambda^2\rho^4\theta^3 + \lambda\rho^4\theta^3 - 2\lambda^3\theta^3 \\ -13\lambda^4\rho^3\theta^3 - 7\beta\lambda^3\rho^3\theta^3 - 6\lambda^2\rho^3\theta^3 + 4\beta\lambda\rho^3\theta^3 + \rho^3\theta^3 + \lambda^5\rho^2\theta^3 + \beta\lambda^4\rho^2\theta^3 + 7\lambda^3\rho^2\theta^3 + \beta\lambda^2\rho^2\theta^3 \\ +2\lambda\rho^2\theta^3 + 2\lambda^4\rho\theta^3 + \beta\lambda^3\rho\theta^3 + \lambda^2\rho\theta^3 - \beta\lambda\rho\theta^3 - \rho\theta^3 - 9\lambda^4\rho^4\theta^2 - 3\beta\lambda^3\rho^4\theta^2 - \lambda^2\rho^4\theta^2 + 17\lambda^3\rho^3\theta^2 \\ +5\beta\lambda^2\rho^3\theta^2 + \lambda\rho^3\theta^2 - 6\lambda^2\rho^2\theta^2 - 2\beta\lambda\rho^2\theta^2 - 2\rho^2\theta^2 - 2\lambda^3\rho\theta^2 + 2\lambda\rho\theta^2 + 7\lambda^3\rho^4\theta + \beta\lambda^2\rho^4\theta + \lambda\rho^4\theta \\ -11\lambda^2\rho^3\theta - \beta\lambda\rho^3\theta - \rho^3\theta + 4\lambda\rho^2\theta - 2\lambda^2\rho^4 + 2\lambda\rho^3 \end{array} \right)}{(1 - \theta\lambda)(1 - \lambda\rho)(\beta\lambda\theta^3 - \lambda^2\theta - \beta\lambda\theta - \theta + 2\lambda)(\rho - \lambda)^2(\lambda\rho\theta - \theta - \rho)^2}$$

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

Consider first the denominator and note that this is a decreasing linear function of  $\beta$ . It is

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

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

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

To verify that the above is necessarily true, we return to condition (A12).

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

$$(A15) \quad \beta < b^* \equiv \frac{\theta^2\lambda^2\rho + \theta^2(-\lambda)\rho^2 + \theta^2\lambda\rho - \theta^2\lambda + \theta^2\rho - \theta\lambda^2\rho - \theta\rho + \lambda\rho}{\theta^3\lambda\rho - \theta^2\lambda\rho^2 - \theta^2\lambda + \theta\lambda\rho}.$$

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

Let  $F(\lambda, \vartheta, \rho)$  denote difference  $\beta^\# - b^*$  as a function of  $(\lambda, \vartheta, \rho)$ ; this function is obtained simply by using the definitions of these thresholds. We have used Mathematica to verify numerically that  $F$  takes non-negative values over the entire  $[0, 1]^3$  set, which itself necessarily contains the admissible values of  $(\lambda, \vartheta, \rho)$ . We conclude that both the numerator and the denominator in (A14) are positive, which means that  $K_{CG}$  is increasing in  $\lambda$  (equivalently, in  $\sigma$ ).

*Proof of Proposition 7*

The proof follows from the main text.

*Proof of Proposition 8*

See Appendix D4.

*Proof of Proposition 9*

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

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

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

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

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

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

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

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

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

where  $\lambda$  is defined in equation (A2). It follows that the average action of group  $g$  is

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

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

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

where

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

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

$$(A16) \quad \Delta_1 = \frac{\rho - \lambda}{\vartheta - \lambda} \Delta_2,$$

$$(A17) \quad b = \left( 1 - \frac{\lambda}{\rho} \right) (\Delta_1 - \Delta_2).$$

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

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

The consumption under perfect information satisfies the standard Euler equation

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

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



Going back to the incomplete-information economy, it follows from (A16) and (A17) that the scale  $b$  is given by

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

and  $\vartheta$  is the inside root of the following equation

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

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

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

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

where

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

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

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

Note that

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

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

#### *Proof of Proposition 10*

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

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

The aggregate outcome for group  $g$  is then

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

By an abuse of notation, we have

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

where  $\bar{\mathbb{E}}_t$  denotes  $[\bar{\mathbb{E}}_{1,t} \dots \bar{\mathbb{E}}_{n,t}]'$ . Denote  $\tilde{\varphi} \equiv (\mathbf{I} - \beta\rho)^{-1}\varphi$  and  $\kappa(L) \equiv (\mathbf{I} - \beta L^{-1})^{-1}\gamma L^{-1}$ . The aggregate outcome  $\mathbf{a}_t$  has the following representation

$$\mathbf{a}_t = \tilde{\varphi} \bar{\mathbb{E}}_t [\xi_t] + \bar{\mathbb{E}}_t [\kappa(L) \tilde{\varphi} \bar{\mathbb{E}}_t [\xi_t]] + \bar{\mathbb{E}}_t [\kappa(L) \bar{\mathbb{E}}_t [\kappa(L) \tilde{\varphi} \bar{\mathbb{E}}_t [\xi_t]]] + \dots$$

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

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

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  $\lambda_g$  is

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

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

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

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  $\mathbf{h}(L)$  is

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

where

$$\mathbf{A}(L) = \text{diag} \left\{ L(L - \lambda_g)(1 - \lambda_g L) \right\} - \beta \text{diag} \left\{ \frac{\lambda_g}{\rho}(L - \rho)(1 - \rho L) + \frac{\lambda_g \tau_g}{\rho} L \right\} - \text{diag} \left\{ \frac{\lambda_g \tau_g}{\rho} L \right\} \gamma,$$

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

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

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

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

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

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

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

$$(A18) \quad \{h_g(0)\}_{g=1}^n, \quad \text{and} \quad \left\{ \beta_g h_g(\lambda_g) \left( \frac{\lambda_g - \rho}{\rho} + \frac{\lambda_g \tau_g}{\rho(1 - \rho \lambda_g)} \right) + \frac{\lambda_g \tau_g}{\rho(1 - \rho \lambda_g)} \sum_k \gamma_{g,k} h_k(\lambda_g) \right\}_{g=1}^n.$$

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

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

The degree of the numerator is  $3n - 1$  as the highest degree of  $\mathbf{d}_g(L)$  is 1 degree less than that of  $A_{g,g}(L)$ . By choosing the constants in equation (A18), the  $2n$  inside roots will be removed. Therefore, the  $2n$  constants are solutions to the following system of linear equations:<sup>2</sup>

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

<sup>2</sup>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.

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

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

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

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

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

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

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

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

Denote  $\tilde{\mathbf{a}}_t \equiv \mathbf{Q} \mathbf{a}_t$ . Because  $\boldsymbol{\Lambda}$  is a diagonal matrix, it follows that  $\tilde{a}_{g,t}$  is independent of  $\tilde{a}_{j,t}$  for  $g \neq j$ , and  $\tilde{a}_{g,t}$  satisfies Proposition 2. The degree of complementarity for  $\tilde{a}_{g,t}$  is  $\mu_g$ , and the corresponding  $\vartheta_g$  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  $\mathbf{a}_t$  is a linear transformation of  $\tilde{\mathbf{a}}_t$ , they share the same AR roots.

*Proof of Proposition 11*

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

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

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

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

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

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

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

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

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

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

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

$$(A19) \quad \mathbf{a}_t = \mathbf{A}\mathbf{a}_{t-1} + \mathbf{B}\xi_t.$$

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

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

If (A19) is a solution to the perfect-information hybrid model, it has to be that

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

By method of undetermined coefficients, we have

$$\begin{aligned} \boldsymbol{\omega}_f\delta(\rho\mathbf{B} + \mathbf{A}\mathbf{B}) &= \mathbf{B} - \boldsymbol{\varphi}, \\ \boldsymbol{\omega}_b &= \mathbf{A}(\mathbf{I} - \boldsymbol{\omega}_f\delta\mathbf{A}). \end{aligned}$$

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,  $\boldsymbol{\omega}_f$  is not uniquely determined.

## B. THE ROLE OF PUBLIC INFORMATION

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

### B1. Solution with a Public Signal

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

$$(B1) \quad z_t = \xi_t + \epsilon_t,$$

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.<sup>3</sup>

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

<sup>3</sup>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.

(i) The equilibrium outcome is given by

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

where  $a_t^\xi$  is the projection of  $a_t$  on the history of  $\xi_t$  and  $v_t$  is the residual.

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

$$(B2) \quad C(z) \equiv -z^3 + \left( \rho + \frac{1}{\rho} + \frac{1}{\rho\sigma^2} + (\delta - \gamma) \right) z^2 - \left( 1 + (\delta - \gamma) \left( \rho + \frac{1}{\rho} \right) + \frac{\delta - \gamma\chi}{\rho\sigma^2} \right) z + (\delta - \gamma).$$

(iii) Provided  $\gamma > 0$ ,  $\vartheta$  is decreasing  $\chi$  and, therefore, both  $\omega_f$  and  $\omega_b$  get closer to their frictionless counterparts as  $\chi$  increases.

(iv) The residual  $v_t$  follows an AR(1) process with innovation  $\epsilon_t$ , the noise in the public signal.

Part (i) expresses the equilibrium outcome as the sum of two components: a “fundamental component,” defined by the projection of  $a_t$  on the history of  $\xi_t$ ; and a residual, itself measurable in the history of  $\epsilon_t$ , the aggregate noise.

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

Part (iii) highlights that, holding  $\sigma$  constant, an increase in  $\chi$  maps to a smaller  $\vartheta$  and, thereby, to smaller distortions, but only if  $\gamma > 0$ ; if instead  $\gamma = 0$ ,  $\chi$  is irrelevant. To understand why, note that an increase in  $\chi$  for given  $\sigma$  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  $\chi$  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  $\chi$  at its lowest possible value? As illustrated next, once the theory is required to match relevant evidence on expectations, the incorporation of public information ( $\chi > 0$ ) may actually translate to *higher* distortions than those predicted by our baseline specification ( $\chi = 0$ ).

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

We find this property is intriguing. If one looks at the response of inflation to either identified monetary shocks (Christiano, Eichenbaum and Evans, 2005; Romer and Romer, 2004) or to the shock that accounts for most of the business cycle volatility in unemployment, output, or the output gap (the MBC shock in Angeletos, Collard and Dellas (2020)), one finds a hump shape. But if one looks at the residual, which the DSGE literature captures with a markup shock, then one sees no hump shape. From this perspective, the introduction of public information helps the theory generate a “residual” in inflation that is of the same type as that found in the data. And it helps reconcile why one sees a hump shape in one dimension but not in another.

#### B2. Revisiting the Mapping from $K_{CG}$ to $(\omega_f, \omega_b)$

Ceteris paribus, the addition of public information reduces the documented distortions by increasing the degree of common knowledge. But it also reduces the predictability of the average

forecasts errors. The relevant question is therefore how the accommodation of public information affects the lessons we draw in this paper under the requirement that the theory continues to match the available evidence on expectations.

In our benchmark, which abstracts from public information, the CG coefficient uniquely identifies the value of  $\sigma$ , which in turn pins down the pair  $(\omega_f, \omega_b)$ , or equivalently the equilibrium dynamics. Now that we have added a public signal, the CG coefficient and the equilibrium dynamics alike depend on two unknown parameters, the precisions  $\tau_x \equiv \sigma_u^{-2}$  and  $\tau_z \equiv \sigma_\epsilon^{-2}$  of, respectively, the private and the public information. As a result, we lose *point* identification but preserve *set* identification: only certain pairs of  $\tau_z$  and  $\tau_x$  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  $\tau_z$ .<sup>4</sup>

Figure B1 illustrates the implications of these properties for the documented distortions within the context of our application to inflation (Section VI). On the horizontal axis, we let  $\tau_z$  vary between zero (our benchmark) and the aforementioned bound. For each  $\tau_z$  in this range, we find the value of  $\tau_x$  that matches the point estimate of  $K_{CG}$  provided in CG and report the implied values for  $\omega_f$  and  $\omega_b$ .

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

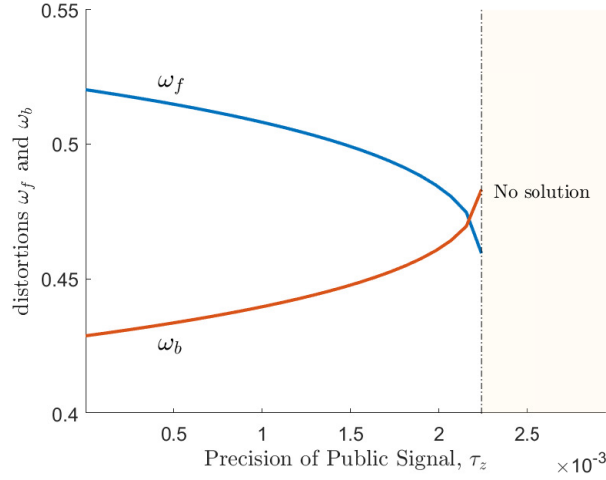


FIGURE B1. THE ROLE OF PUBLIC INFORMATION

Similar points apply if we let for an *endogenous* public signal of the form  $z_t = a_t + \epsilon_t$ , which in the application under consideration can be thought of as statistic of inflation contaminated with measurement error.<sup>5</sup> Similar to the exogenous-information case, matching the CG moment puts an upper bound on the informativeness of this signal. Different from the exogenous-information case, this informativeness is now endogenous to the actual inflation dynamics. This introduces an

<sup>4</sup>That is, the set of the admissible values for the pair  $(\tau_x, \tau_z)$  can be expressed as

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

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

<sup>5</sup>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).

additional fixed point problem, which can only be solved numerically. But as illustrated in Figure D1 in Appendix D3, the main message goes through.

### C. IDIOSYNCRATIC SHOCKS AND MICRO- VS MACRO-LEVEL DISTORTIONS

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

Consider next the menu-cost literature that aims at accounting for the microeconomic data on prices (Goloso 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 Maćkowiak and Wiederholt (2009). It follows that the documented distortions may loom large at the macroeconomic time series even if they appear to be small in the microeconomic time series.

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

$$(C1) \quad a_{i,t} = \mathbb{E}_{i,t}[\varphi \xi_{i,t} + \beta a_{i,t+1} + \gamma a_{t+1}],$$

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  $t$ .<sup>6</sup>

We then specify the information structure as follows. First, we let each agent observe the same signal  $x_{i,t}$  about the aggregate shock  $\xi_t$  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  $\xi_t$ , 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

<sup>6</sup>The restriction that the two kinds of shocks have the same persistence is only for expositional simplicity.



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 III.C to the current specification. It can thus be shown that the equilibrium action is given by the following:

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

where  $\vartheta$  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  $\xi_t$  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  $\vartheta$  and  $\lambda$  increases with the strength of the GE effect, as measured with  $\gamma$ .

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,

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

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  $\lambda$ , 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  $\vartheta$  remains higher than  $\hat{\lambda}$  even if  $\hat{\lambda} = \lambda$ . That is, even if the first-order uncertainty about the two kind of shocks is the same, the distortion at the macro level may remain larger insofar as there are positive GE feedback effects, such as the Keynesian income-spending multiplier or the dynamic strategic complementarity in price-setting decisions of the firms.

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

## D. APPLICATION TO INFLATION: MICRO-FOUNDATIONS AND ADDITIONAL RESULTS

### D1. 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:

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

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  $\mathbf{mc}_{t+k}$  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:

$$(D1) \quad p_{i,t}^* = (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,t} [\mathbf{mc}_{t+k} + p_{t+k}].$$

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 VI](#), 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 [\(D1\)](#) as

$$(D2) \quad p_{i,t}^* - p_{t-1} = (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,t} [\mathbf{mc}_{t+k}] + \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,t} [\pi_{t+k}],$$

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

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

Combining this with condition (D2) and rearranging, we arrive at the following expression:

$$(D3) \quad \pi_t = \kappa \sum_{k=0}^{\infty} (\chi\theta)^k \bar{\mathbb{E}}_t [\mathbf{mc}_{t+k}] + \chi(1-\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \bar{\mathbb{E}}_t [\pi_{t+k+1}].$$

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

When information is complete, we can replace  $\bar{\mathbb{E}}_t[\cdot]$  with  $\mathbb{E}_t[\cdot]$ , the expectation of the representative agent. We can then use the Law of Iterated Expectations to reduce condition (D3) 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 (D3) cannot be reduced to the standard NKPC.

As explained in the main text, condition (D3) 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 (D3) 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 (D1) and rewrite it in recursive form as follows:

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

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:

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

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.

## D2. 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 III.A.

Using condition (D3), the incomplete-information inflation dynamics can be decomposed into two components: the belief of the present discounted value of real marginal costs,  $\varphi \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[\mathbf{mc}_{t+k}]$ ; and the belief of of the present discounted value of inflation,  $\gamma \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[\pi_{t+k+1}]$ . The same

decomposition can also be applied when agents have perfect information:

$$(D4) \quad \pi_t^* = \underbrace{\varphi \sum_{k=0}^{\infty} \beta^k \mathbb{E}_t [\mathbf{mc}_{t+k} | \mathbf{mc}_t]}_{\text{complete-info PE component}} + \underbrace{\gamma \sum_{k=0}^{\infty} \beta^k \mathbb{E}_t [\pi_{t+k+1}^* | \mathbf{mc}_t]}_{\text{complete-info GE component}}.$$

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:

$$(D5) \quad \tilde{\pi}_t = \underbrace{\varphi \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t [\mathbf{mc}_{t+k}]}_{\text{incomplete-info PE component}} + \underbrace{\gamma \sum_{k=0}^{\infty} \beta^k \mathbb{E}_t [\pi_{t+k+1}^* | \mathbf{mc}_t]}_{\text{complete-info GE component}}.$$

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

The logic behind this finding can be understood by computing the GE multiplier that is hidden inside the standard NKPC. Let  $\mu^*$  be the ratio of the GE component to the PE component under complete information, that is, the ratio of the two terms seen in condition (D4). This identifies the GE multiplier; the total effect is  $1 + \mu^*$  times the PE effect. Straightforward calculation shows that

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

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

### D3. Adding Public Information

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

We thus consider two cases: an *exogenous* public signal of the form  $z_t = \mathbf{mc}_t + \text{noise}$ , and an *endogenous* public signal of the form  $z_t = \pi_t + \text{noise}$ , namely a noisy statistic of inflation. The first case affords an analytical characterization, along the lines of Appendix B; the second case requires a numerical approximation but, as shown below, only reinforces our message.<sup>7</sup>

Figure D1 compares the IRF of inflation to innovations in the real marginal cost under three information structures, all required to match the regression coefficient  $K_{CG}$  estimated in CG. The blue, solid line corresponds to our benchmark, which abstracts from public information. As explained in Appendix B, once we allow for a public signal, there is a range of admissible values for its precision, each one mapping to a different pair  $(\omega_f, \omega_b)$ , or a different IRF. The red, dashed line in the figure gives the IRF that is obtained when the public signal is exogenous and its precision is the maximal one consistent with  $K_{CG}$ . The area between this line and the benchmark line spans all the admissible parameterizations of the exogenous-information case. Finally, the black, dotted line

<sup>7</sup>We thank an anonymous referee for suggesting these explorations.

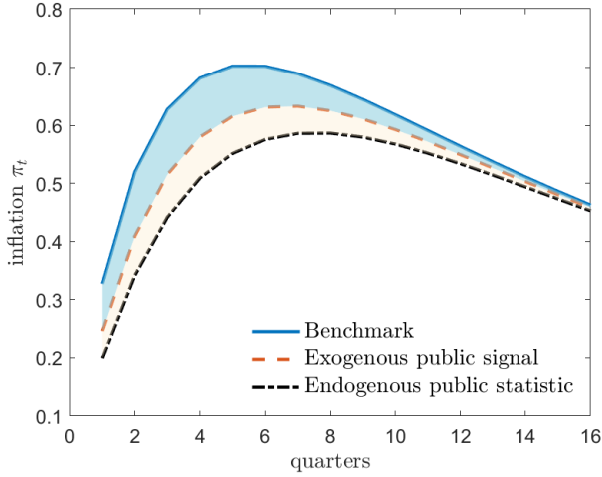


FIGURE D1. IRF OF INFLATION, EXOGENOUS VS ENDOGENOUS INFORMATION

gives the IRF that obtains when the public signal is endogenous and its precision equals the appropriate upper bound. The area between this line and the benchmark line spans all the admissible parameterizations of the endogenous-information case.

The main takeaways are twofold. First, the exogenous-information setting provides a useful analytical tool to understand the more realistic but less tractable endogenous-information case. Second, the accommodation of public information, exogenous or endogenous, only reinforces the quantitative findings once the theory is disciplined by the available evidence on expectations.<sup>8</sup>

#### D4. Market Concentration

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

LEMMA D.1: *The optimal reset price of individual firm  $i$  in market  $m$  follows*

$$(D6) \quad p_{i,m,t}^* = (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \kappa \mathbb{E}_{i,m,t}[mc_{t+k}] + (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,m,t}[\alpha_N p_{m-i,t+k} + (1 - \alpha_N)p_{t+k}],$$

where  $\alpha_N$  is given by

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

In condition (D6),  $\chi$ ,  $\theta$ , and  $\kappa$  are the same parameters as in the baseline NKPC setup, while  $\alpha_N \in (0, 1)$  is a new scalar which summarizes how much a firm's pricing strategy depends on the prices of its competitors relative to the aggregate price level. It is easy to verify that  $\psi > 1$  and  $\psi > \epsilon$  suffices for  $\alpha_N$  to be decreasing in  $N$ . And in the special case in which  $\psi = \infty$ , which amounts to a Cournot-like game for each market, we have more simply that  $\alpha_N = 1/(2N)$ .

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

<sup>8</sup>A third, subtler takeaway is that the endogenous public signal contributes to more persistence than the exogenous one. We find this intriguing and we suspect it is because inflation moves more sluggishly than the fundamental, thus slowing down the learning. Nimark (2008) also hypothesizes that endogenous signals add persistence. The logic is, however, complicated by the fact that, as we vary the form of the signal, we adjust its precision to make sure that theory keeps matching the CG moment.

LEMMA D.2: *The aggregate inflation rate follows*

(D7)

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

For our purposes, the key observation is that  $\alpha_N$  is decreasing in  $N$ , or decreasing in market concentration. Intuitively, as  $N \rightarrow \infty$  and the firm becomes infinitesimally small not only vis-a-vis the entire economy but also vis-a-vis its own market, the firm only care to set a price in proportion to its nominal marginal cost, which itself is driven by the aggregate price level. That is, as  $N \rightarrow \infty$ ,  $\alpha_N$  approaches 1, condition (D7) reduces to condition (25), and we recover the case studied before. But when  $N$  is finite, a new consideration emerges: when a firm raises its price, it depresses its market share. This effect scales up with market concentration, explaining why higher market concentration maps to a higher  $\alpha_N$ , or a higher consideration for local conditions relative to aggregate conditions.

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

These points are evident from condition (D7). Mapping this condition to our framework yields

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

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

## E. HETEROGENEITY À LA HANK

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

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

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

where  $C_{i,g,\tau;t}$  denotes her consumption in period  $t$  (conditional on survival) and  $\chi \in (0, 1)$  is the

subjective discount factor. Her budget constraint, on the other hand, is given by

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

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

We henceforth work with the log-linearized solution around a steady state in which there are no shocks,  $\chi R_t = 1$ , and  $C_t = Y_t = Y^*$ , where  $Y^*$  is the natural rate of output.<sup>9</sup> We use lower-case variables to represent log-deviations from the steady state (e.g.,  $r_t \equiv \log R_t - \log \chi^{-1}$ ), with the exception that  $s_{i,g,\tau;t}$  and  $\tau_{g,t}$  stand for, respectively,  $\frac{S_{i,g,\tau;t}}{Y^*}$  and  $\frac{T_{g,t}}{Y^*}$  as their steady-state values are zero. We can then express the optimal expenditure of a consumer in group  $g$  as follows:

$$(E1) \quad c_{i,g,\tau,t} = (1 - \chi \varpi_g) \left( \frac{1}{\chi \varpi_g} s_{i,g,\tau,t-1} + \mathbb{E}_{i,t}[\mathcal{T}_{g,t}] \right) - \chi \omega_g \sum_{j=0}^{\infty} (\chi \varpi_g)^j \mathbb{E}_{i,t}[r_{t+j}] \\ + (1 - \chi \varpi_g) \phi_g \sum_{j=0}^{\infty} (\chi \varpi_g)^j \mathbb{E}_{i,t}[y_{t+j}]$$

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

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

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

Aggregating (E1) across all consumers of any given group  $g$ , we get

$$(E2) \quad c_{g,t} = (1 - \chi \omega_g) \left( \frac{1}{\chi} s_{g,t-1} + \bar{\mathbb{E}}_t[\mathcal{T}_{g,t}] \right) - \chi \omega_g \sum_{j=0}^{\infty} (\chi \omega_g)^j \bar{\mathbb{E}}_t[r_{t+j}] \\ + (1 - \chi \omega_g) \phi_g \sum_{j=0}^{\infty} (\chi \omega_g)^j \bar{\mathbb{E}}_t[y_{t+j}].$$

Similarly, by aggregating the budget constraints of all consumers in group  $g$ , and taking into account how the annuities effectively redistribute wealth from deceased to surviving agents, we get the following group-level budget constraint:

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

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

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

<sup>9</sup>To simplify the exposition, we suppress the production side of the economy and the determination of the flexible-price outcomes. The details can be filled in the usual way; let technology be linear in labor and assume constant aggregate productivity to get a time-invariant natural rate of output.

interpreted monetary shocks.

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

$$(E3) \quad \sum_g \pi_g (1 - \chi \omega_g) s_{g,t} + \sum_g \pi_g \bar{\mathbb{E}}_t[\mathcal{T}_{g,t}] = 0,$$

When all groups have the same MPC (i.e.,  $\omega_g = \omega_{g'}$  for all  $g, g'$ ), this condition is trivially satisfied with  $\mathcal{T}_{g,t} = 0$  for all  $g, t$ . When instead different groups have different MPCs, this condition requires that fiscal policy offsets the interaction of MPC heterogeneity with wealth inequality. In particular, a sufficient condition for (E3) to hold is that  $\bar{\mathbb{E}}_t[\mathcal{T}_{g,t}] = (1 - \chi \omega_g) s_{g,t}$  for all  $g, t$ . And since  $s_{g,t}$  is measurable in the history of the aggregate shock alone, the transfers do *not* have to be conditioned on the consumers' age or idiosyncratic histories.

As long as condition (E3) is satisfied, we can aggregate condition (E2) across groups to obtain the economy-wide aggregate consumption as follows:

$$(E4) \quad c_t = \sum_g \pi_g \left\{ -\chi \omega_g \sum_{j=0}^{\infty} (\chi \omega_g)^j \bar{\mathbb{E}}_t[r_{t+j}] + (1 - \chi \omega_g) \phi_g \sum_{j=0}^{\infty} (\chi \omega_g)^j \bar{\mathbb{E}}_t[y_{t+j}] \right\}$$

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

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

and where  $y_t = \sum_g \pi_g y_{t,g}$ . Note that  $c_{g,t}$ , the actual consumption of group  $g$ , may differ from  $y_{g,t}$ , the auxiliary variable introduced above. This will indeed be the case whenever  $\bar{\mathbb{E}}_t[\mathcal{T}_{g,t}] \neq (1 - \chi \omega_g) s_{g,t}$  for some  $g$  and some  $t$ . Still, as long as (E3) is satisfied, the economy-wide outcomes are determined in the manner described above—and coincide with those reported in the main text.

This completes the details behind Figure 4. Consider next what happens when condition (E3) is violated and, as a result, wealth inequality can feed into the aggregate dynamics. In particular, impose  $\mathcal{T}_{g,t} = 0$  for all  $g, t$ . If all groups had the same MPC, (E3) and (E4) would still hold; but then the heterogeneity in business-cycle exposure would also not matter. The interesting case is when fiscal policy is inactive and, in addition, there is joint heterogeneity in the business-cycle exposure and the MPC. This case is studied in Figure 5 in the main text.

## F: APPLICATION TO INVESTMENT

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

$$(F1) \quad \text{Cost}_t = \Phi \left( \frac{I_t}{K_{t-1}} \right)$$

where  $I_t$  denotes the rate of investment,  $K_{t-1}$  denotes the capital stock inherited from the previous period, and  $\Phi$  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).<sup>10</sup>

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:

$$(F2) \quad \text{Cost}_t = \Psi \left( \frac{I_t}{I_{t-1}} \right)$$

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 (F1); 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 (F2). 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

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

subject to

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

Here, the fundamental shock,  $\xi_t$ , 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

<sup>10</sup>These works differ on the importance they attribute to heterogeneity, lumpiness, and non-linearities, but appear to share the prediction that the impulse response of aggregate investment is peaked on impact. They therefore do not provide a micro-foundation of the kind of sluggish investment dynamics featured in the DSGE literature.

of the final goods producer's optimal condition around the balanced growth path can be written as

$$(F3) \quad Qq_t + \psi Z z_t = \chi \mathbb{E}_t \left[ A\xi_{t+1} + Qq_{t+1} + \psi Z(1+Z)z_{t+1} \right].$$

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

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

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

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

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  $\xi_t$  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:

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

where

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

**Incomplete Information.** Suppose now that firms receive a noisy signal about the fundamental  $\xi_t$  as in Section II. 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 (F3), the aggregate investment dynamics follow

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

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

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

It is then immediate that when information is incomplete, there exist  $\omega_f < 1$  and  $\omega_b > 0$  such that

the equilibrium process for investment solves the following equation:

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

Finally, it is 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:

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

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

## G: APPLICATION TO ASSET PRICES

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

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

where  $p_t$  is the price of the asset in period  $t$ ,  $d_{t+1}$  is its dividend in the next period,  $\mathbb{E}_t$  is the expectation of the representative agent, and  $\chi$  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:

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

where  $\bar{\mathbb{E}}_t$  is the *average* expectation of the traders in period  $t$  and  $\epsilon_t$  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  $\xi_t$  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).<sup>11</sup> 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  $\xi_t$  obeys the following law of motion, for some  $\omega_f < 1$  and  $\omega_b > 0$ :

$$(G1) \quad p_t = \mathbb{E}_t[d_{t+1}] + \omega_f \chi \mathbb{E}_t[p_{t+1}] + \omega_b p_{t-1},$$

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.

<sup>11</sup>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.

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 (G1) 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 VI, the quantitative implications may be similar provided that the theory is disciplined with the relevant evidence on expectations.

#### H. ROBUSTNESS OF MAIN INSIGHTS

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

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

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

**Setup.** We henceforth let the fundamental  $\xi_t$  follow a flexible, possibly infinite-order, MA process:

$$(H1) \quad \xi_t = \sum_{k=0}^{\infty} \rho_k \eta_{t-k},$$

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  $\eta_{t-k}$  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  $\tau_k$  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.<sup>12</sup> 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 III](#) because a single parameter,  $\sigma$ , controlled both objects at once.

**Dynamics of Higher-Order Beliefs.** The information regarding  $\eta_{t-k}$  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  $\bar{\mathbb{E}}_t[\eta_{t-k}] = \lambda_k \eta_{t-k}$  and therefore

$$(H2) \quad \bar{\mathbb{E}}_t[\xi_t] = \bar{\mathbb{E}}_t \left[ \sum_{k=0}^{\infty} \rho_k \eta_{t-k} \right] = \sum_{k=0}^{\infty} f_{1,k} \eta_{t-k}, \quad \text{with} \quad f_{1,k} = \lambda_k \rho_k.$$

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  $\lambda_0$  is (i.e., the less precise the initial information is), the larger the initial gap between the two IRFs (i.e., a larger the initial forecast error). And the slower  $\lambda_k$  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 III](#), 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 [\(H2\)](#) to period  $t + 1$

<sup>12</sup>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.

and taking the period- $t$  average expectation, we get

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

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

$$(H3) \quad \bar{\mathbb{F}}_t^h [\xi_{t+h-1}] = \sum_{k=0}^{\infty} f_{h,k} \eta_{t-k}, \quad \text{with} \quad f_{h,k} = \lambda_k \lambda_{k+1} \dots \lambda_{k+h-1} \rho_{k+h-1}.$$

The increasing components in the product  $\lambda_k \lambda_{k+1} \dots \lambda_{k+h-1}$  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

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

The following result is thus immediate.

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

- (i) *For all  $k$  and all  $h$ , this ratio is strictly between 0 and 1.*
- (ii) *For any  $k$ , this is decreasing in  $h$ .*
- (iii) *For any  $h$ , this ratio is increasing in  $k$ .*
- (iv) *As  $k \rightarrow \infty$ , 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  $\lambda_k \rightarrow 1$ .*

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  $k \rightarrow \infty$  if and only if  $\lambda_k \rightarrow 1$ , 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 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  $\eta_t$  in some period  $t$ . A fraction  $1 - q$  of the population becomes aware of it immediately and hence  $\bar{\mathbb{E}}_t[\eta_t] = (1 - q)\eta_t$ . A period later, an additional  $(1 - q)q$  fraction becomes aware of it and hence  $\bar{\mathbb{E}}_{t+1}[\eta_t] = (1 - q^2)\eta_t$ . And so on. It follows that sticky information 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  $a_t = \bar{\mathbb{E}}_t[\xi_t] + \gamma \bar{\mathbb{E}}_t[a_{t+1}]$ , which in turn implies that  $a_t = \sum_{h=1}^\infty \gamma^{h-1} \bar{\mathbb{F}}_t^h[\xi_{t+h-1}]$ . From the preceding characterization of the higher-order beliefs  $\bar{\mathbb{F}}_t^h[\xi_{t+h-1}]$ , it follows that

$$(H4) \quad a_t = \sum_{k=0}^\infty g_k \eta_{t-k}, \quad \text{with} \quad g_k = \sum_{h=1}^\infty \gamma^{h-1} f_{h,k} = \left\{ \sum_{h=1}^\infty \gamma^{h-1} \lambda_k \lambda_{k+1} \dots \lambda_{k+h-1} \rho_{k+h-1} \right\}.$$

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  $\gamma$  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.<sup>13</sup>

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

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

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

$$(H5) \quad a_t = \xi'_t + \gamma' \mathbb{E}_t[a_{t+1}],$$

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  $\xi'_t$  with  $\xi_t$  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 III.E.

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-

<sup>13</sup>The last point is particularly clear if we set  $\rho_k = \rho^k$  (meaning that  $\xi_t$  follows an AR(1) process). In this case, the initial response is given by

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

from which it is evident that the importance of higher-order beliefs increases with both  $\gamma$  and  $\rho$ . This further illustrates the point made in Section III.D regarding the role of the persistence of the fundamental.

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  $\xi_t$ , 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  $\xi_t$  and is given by

$$a_t = \sum_{k=0}^{\infty} \widehat{g}_k \eta_{t-k}, \quad \text{with} \quad \widehat{g}_k = \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k \rho_{k+h-1}.$$

This resembles the complete-information benchmark in that the outcome is pinned down by the first-order belief of  $\xi_t$ , but allows this belief to adjust sluggishly to the underlying innovations in  $\xi_t$ .

By construction, the variant economy preserves the effects of learning on first-order beliefs but shuts down the interaction of learning with higher-order uncertainty. It follows that the comparison of this economy with the original economy reveals the role of this interaction.

**PROPOSITION 14:** *Let  $\{g_k\}$  and  $\{\widehat{g}_k\}$  denote the Impulse Response Function of the equilibrium outcome in the two economies described above.*

(i)  $0 < g_k < \widehat{g}_k$  for all  $k \geq 0$

(ii) If  $\frac{\rho_k}{\rho_{k-1}} \geq \frac{\rho_{k+1}}{\rho_k}$  and  $\rho_k > 0$  for all  $k > 0$ , then  $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$  for all  $k \geq 0$

Consider property (i), in particular the property that  $g_k < \widehat{g}_k$ . This property means that our economy exhibits a uniformly smaller dynamic response for the equilibrium outcome than the aforementioned economy, in which higher-order uncertainty is shut down. But note that the two economies share the following law of motion:

$$(H6) \quad a_t = \varphi \overline{\mathbb{E}}_t[\xi_t] + \gamma \overline{\mathbb{E}}_t[a_{t+1}].$$

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  $\xi_t$ ) 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:

$$(H7) \quad g_k = f_{1,k} + \lambda_k \gamma g_{k+1}.$$

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  $\rho_k$ . The role of higher-order uncertainty is captured by the presence of  $\lambda_k$  in the second term: it is as if the discount factor  $\gamma$  has been replaced by a discount factor equal to  $\lambda_k \gamma$ , which is strictly less than  $\gamma$ . This represents a generalization of the form of myopia seen in condition (H5). There, learning was shut down, so that that  $\lambda_k$  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  $\lambda_k$  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.<sup>14</sup>

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,<sup>15</sup> 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 1:** *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  $\lambda_k$ ,  $\widehat{g}_k$ , 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  $\omega_b$ , which captures the endogenous build up in the equilibrium dynamics, is positive because of learning and it is higher the higher the importance of higher-order uncertainty.

**Multiple Fundamental Shocks.** So far, we have focused on the case where there is a single fundamental shock. Now we extend the analysis to a case where multiple fundamental shocks are present. On one hand, we will show that relative to the frictionless benchmark, when these shocks cannot be perfectly separated, agents may overact to some of these shocks and underact to the others when we focus on the PE effects, as in Lucas (1976). On the other hand, we will show that higher-order uncertainty, which exclusively related to the GE effects, still results in distortions in the form of myopia and anchoring *relative* to its complete-information counterpart.

Suppose that the best response is

$$a_{i,t} = \mathbb{E}_{i,t}[\phi_1 \xi_t^1 + \phi_2 \xi_t^2] + \gamma \mathbb{E}_{i,t}[a_{t+1}],$$

<sup>14</sup>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.

<sup>15</sup>Recall, this is by property (i) of Proposition 14.

where the two fundamental shocks are driven by two different innovations  $\eta_t^1$  and  $\eta_t^2$

$$\xi_t^1 = \sum_{k=0}^{\infty} \rho_k^1 \eta_{t-k}^1, \quad \text{and} \quad \xi_t^2 = \sum_{k=0}^{\infty} \rho_k^2 \eta_{t-k}^2.$$

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

$$\bar{\mathbb{E}}_t[\eta_{t-k}] = \lambda_k.$$

where the sequence of  $\lambda_k$  is defined in a similar way as before. The average expectations on each of the aggregate innovations is given by

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

where the weights  $\phi_1$  and  $\phi_2$  depend on the relative volatility of  $\eta_t^1$  versus  $\eta_t^2$ , 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

$$\begin{aligned} \bar{\mathbb{E}}_t[\phi_1 \xi_t^1] &= \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k} = \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k}^1 + \phi_1 \varpi_1 \sum_{k=0}^{\infty} \lambda_k \rho_k^1 \eta_{t-k}^2, \\ \bar{\mathbb{E}}_t[\phi_2 \xi_t^1] &= \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k} = \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k}^1 + \phi_2 \varpi_2 \sum_{k=0}^{\infty} \lambda_k \rho_k^2 \eta_{t-k}^2. \end{aligned}$$

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,  $\eta_t^1$ . In the frictionless case,  $\bar{\mathbb{E}}_t[\omega_1 \xi_t^1] = \omega_1 \sum_{k=0}^{\infty} \rho_k^1 \eta_{t-k}^1$ . The average expectation of  $\xi_t^1$  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  $\eta_t^2$  due to informational frictions. The total effects could well be a higher response overall.

Now we turn to the effects of the GE consideration and higher-order uncertainty with  $\gamma > 0$ . The average higher-order expectations are given by

$$\bar{\mathbb{F}}_t^h[\omega_1 \xi_{t+h-1}^1 + \omega_2 \xi_{t+h-1}^2] = \sum_{k=0}^{\infty} f_{h,k} \eta_{t-k}, \quad \text{with} \quad f_{h,k} = \lambda_k \lambda_{k+1} \dots \lambda_{k+h-1} (\omega_1 \phi_1 \rho_{k+h-1}^1 + \omega_2 \phi_2 \rho_{k+h-1}^2).$$

Here, we utilize the property that agents cannot separate  $\eta_t^1$  from  $\eta_t^2$  and the expectations can be effectively written as functions of  $\eta_t$ .

Similar to the single-shock economy, the aggregate outcome can be written as (H8)

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

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

aggregate outcome is

$$(H9) \quad a_t = \sum_{k=0}^{\infty} \hat{g}_k \eta_{t-k}, \quad \text{with} \quad \hat{g}_k = \left\{ \sum_{h=1}^{\infty} \gamma^{h-1} \lambda_k (\omega_1 \phi_1 \rho_{k+h-1}^1 + \omega_2 \phi_2 \rho_{k+h-1}^2) \right\}.$$

Define  $\hat{\xi}_t$  as

$$\hat{\xi}_t \equiv \sum_{k=0}^{\infty} (\omega_1 \phi_1 \rho_k^1 + \omega_2 \phi_2 \rho_k^2) \eta_{t-k}.$$

By replacing  $\xi_t$  by  $\hat{\xi}_t$ , the analysis on myopia and anchoring in Proposition 14 extends to the current setting. Therefore, relative to the complete-information counterpart, the effects of additional myopia and anchoring remain the same when there exist multiple fundamental shocks.

**Two Forms of Bounded Rationality.** We now shed light on two additional points, which were anticipated earlier on: the role played by the anticipation that others will learn in the future; and the possible interaction of incomplete information with Level-k Thinking.

To illustrate the first point, we consider a behavioral variant where agents fail to anticipate that others will learn in the future. To simplify, we also set  $\beta = 0$ . Recall from equation (H3), when agents are rational, the forward higher-order beliefs are

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

In the variant economy, by shutting down the anticipation of learning, the nature of higher-order beliefs changes, as  $\bar{\mathbb{E}}_{i,t} [\bar{\mathbb{E}}_{t+k} [\xi_{t+q}]] = \bar{\mathbb{E}}_{i,t} [\bar{\mathbb{E}}_t [\xi_{t+q}]]$  for  $k, q \geq 0$ , and the counterpart of  $\bar{\mathbb{F}}_t^h [\xi_{t+h-1}]$  becomes

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

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

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

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:  $\omega_f$  is smaller and  $\omega_b$  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  $a_t^0 = \bar{\mathbb{E}}_t[\xi_t]$ .

With level-1 thinking, agent  $i$ 's action changes to

$$a_{i,t}^1 = \mathbb{E}_{i,t}[\xi_t] + \gamma \mathbb{E}_{i,t}[a_{t+1}^0] = \mathbb{E}_{i,t}[\xi_t] + \gamma \mathbb{E}_{i,t} [\bar{\mathbb{E}}_{t+1}[\xi_{t+1}]],$$

where the second-order higher-order belief shows up. By induction, the level- $k$  outcome is given by

$$a_t^k = \sum_{h=1}^{k+1} \gamma^{h-1} \bar{\mathbb{F}}_t^h [\xi_{t+h-1}].$$

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.<sup>16</sup> That said, the source of the additional myopia is different. In the first, the relevant forward-looking higher-order beliefs have been replaced by myopic counterparts, which move less. In the second, the right, forward-looking higher-order beliefs are still at work, but they have been truncated at a finite point.

## I. MULTIPLE SHOCKS

Our baseline specification has assumed that there is a single shock that drives the fundamental. In this section, we extend our analysis in the direction of [Kohlhas and Walther \(2019\)](#) to include both procyclical and countercyclical components, and show that a modified version of our main result holds.

Consider the following best response, which is similar to our baseline specification:

$$(I1) \quad y_{it} = \varphi \mathbb{E}_{it}[\zeta_t] + \beta \mathbb{E}_{it}[y_{it+1}] + \gamma \mathbb{E}_{it}[y_{t+1}].$$

But now allow the fundamental  $\zeta_t$  to be driven by  $N$  different components:

$$\zeta_t = \sum_{j=1}^N d_{jt}, \quad \text{with} \quad d_{jt} = \kappa_j \xi_t + \epsilon_{j,t}.$$

The common shock among different components,  $\xi_t$ , follows an AR(1) process:

$$\xi_t = \rho \xi_{t-1} + \eta_t.$$

The component-specific shocks  $\epsilon_{j,t} \sim \mathcal{N}(0, \tau_j^{-1})$  are i.i.d. across both  $j$  and  $t$ . The loading of component  $j$  on  $\xi_t$  is  $\kappa_j$ , which could be both positive or negative, capturing for procyclical or counter-cyclical components. Finally,  $\sum_j \kappa_j = 1$ .

In terms of the information structure, assume that each agent receives  $N$  private signals, one per component:

$$x_{i,j,t} = d_{jt} + u_{i,j,t}, \quad u_{i,j,t} \sim \mathcal{N}(0, \omega_j^{-1}).$$

This is the same structure considered in [Kohlhas and Walther \(2019\)](#), which leads to asymmetric attention by allowing heterogeneity in  $\omega_j$ .

To see how this structure connects with our equivalence result, we turn to the following auxiliary best response in which only the persistent shock  $\xi_t$  is pay-off relevant:

$$(I2) \quad a_{it} = \varphi \mathbb{E}_{it}[\xi_t] + \beta \mathbb{E}_{it}[a_{it+1}] + \gamma \mathbb{E}_{it}[a_{t+1}].$$

This best response is exactly the same as our model. The aggregate outcome  $y_t$  from condition (I1) is related to the aggregate outcome  $a_t$  from condition (I2) in the following way:

$$\begin{aligned} y_t &= \varphi \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[\zeta_{t+k}] + \gamma \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[y_{t+k+1}] \\ &= \varphi \sum_{j=1}^N \bar{\mathbb{E}}_t[\epsilon_{j,t}] + \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[\xi_{t+k}] + \gamma \sum_{k=0}^{\infty} \beta^k \bar{\mathbb{E}}_t[y_{t+k+1}] \end{aligned}$$

<sup>16</sup>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- $\infty$  outcome in the second variant.

$$= \varphi \sum_{j=1}^N \bar{\mathbb{E}}_t[\epsilon_{j,t}] + a_t,$$

where the last equality is due to that only the persistent shock  $\xi_t$  matters for  $y_{t+k}$  in the future, and the forecasts of the transitory shocks  $\epsilon_{j,t}$  are zero. We conclude that

$$y_t = a_t + u_t,$$

where  $u_t \equiv \varphi \sum_{j=1}^N \bar{\mathbb{E}}_t[\epsilon_{j,t}]$ .

Consider how  $u_t$  is determined. To this goal, let us first compute the forecast of the persistent shock  $\xi_t$ . Since this object only involves the first-order belief, it is more convenient to consolidate the  $N$  different signals into a single one

$$x_{i,t} = \xi_t + \frac{1}{\tau} \sum_{j=1}^N \kappa_j (\tau_j^{-1} + \omega_j^{-1})^{-1} (\epsilon_{j,t} + u_{i,j,t}) \equiv \xi_t + u_{i,t}, \quad u_{i,t} \sim \mathcal{N}(0, \tau^{-1}),$$

where  $\tau = \sum_{j=1}^N \kappa_j^2 (\tau_j^{-1} + \omega_j^{-1})^{-1}$ . That is, it is as if each agent observes a single signal, which however contains both idiosyncratic and aggregate noise—a hybrid of the private and public signals considered in Appendix B. Using this observation, we can compute the average forecast as follows:

$$\bar{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{g}{\rho}\right) \frac{1}{1 - gL} \sum_{j=1}^N \frac{\kappa_j^2 (\tau_j^{-1} + \omega_j^{-1})^{-1}}{\tau} (\xi_t + \kappa_j^{-1} \epsilon_{j,t}),$$

or equivalently

$$(I3) \quad \bar{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{g}{\rho}\right) \frac{1}{1 - gL} (\xi_t + \epsilon_t),$$

where  $g \equiv \frac{1}{2} \left[ \rho + \frac{1}{\rho} (1 + \tau) - \sqrt{\left(\rho + \frac{1}{\rho} (1 + \tau)\right)^2 - 4} \right] < \rho$  and  $\epsilon_t \equiv \frac{1}{\tau} \sum_{j=1}^N \kappa_j (\tau_j^{-1} + \omega_j^{-1})^{-1} \epsilon_{j,t}$ .

Next, denote with  $\lambda_j \equiv \frac{\omega_j}{\tau_j + \omega_j} \in (0, 1)$  the signal-to-noise ratio applied when inferring  $\epsilon_{j,t}$  from  $u_{i,j,t}$ . The average forecast of the sum of component-specific shocks is given by

$$(I4) \quad \sum_{j=1}^N \bar{\mathbb{E}}_t[\epsilon_{j,t}] = \sum_{j=1}^N \lambda_j (\kappa_j \xi_t + \epsilon_{j,t} - \kappa_j \bar{\mathbb{E}}_t[\xi_t]) = \sum_{j=1}^N \lambda_j \kappa_j (\xi_t - \bar{\mathbb{E}}_t[\xi_t]) + \sum_{j=1}^N \lambda_j \epsilon_{j,t}.$$

It follows that the determination of  $u_t$  boils down to a pure forecasting problem spelled out by equations (I3) and (I4).

Consider next the determination of  $a_t$ . As already mentioned, this obtains from the same best responses as our model, with  $\xi_t$  been the sole fundamental. The information structure about it is more complicated than in our baseline analysis, as agents observed signals contaminated with both idiosyncratic and common noise. But a result similar to Proposition 12 in Appendix B applies. That is,

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

where  $a_t^\xi$ , the fundamental component, obeys our observational equivalence result and  $\nu_t$ , the residual, is an AR(1) driven by the “noise” (here, the combination of the  $\epsilon_{j,t}$ ’s). The only subtle difference is in the precise cubic that pins down  $\vartheta$  (and thereby  $\omega_f$  and  $\omega_b$ ).

To complete the picture, consider the projection of  $y_t$  on the history of  $\xi_t$ . This is given by

$$y_t^\xi = \frac{\tilde{\varphi}}{1 - gL} \eta_t + a_t^\xi,$$

where  $\eta_t$  is the innovation in  $\xi_t$  and  $\tilde{\varphi} \equiv \varphi \sum_{j=1}^N \lambda_j \kappa_j \frac{q}{\rho}$ . We thus have that the IRF of  $y_t$  with respect to  $\eta_t$  is the sum of the AR(2) corresponding to  $a_t^x$  and of the AR(1) given by the first term above. Clearly, this term does not contribute to a hump-shape. Furthermore, it is likely to be quantitatively less important than  $a_t^\xi$  for the following reason:  $a_t^\xi$  consists of all the PE and GE effects across all the horizons, while  $\frac{\tilde{\varphi}}{1 - gL} \eta_t$  captures only a fraction of the total PE effects. For instance, as explained in Appendix D2, in our inflation applications GE effects are about 7 times as large as PE effects. This suggests that, in that context, the  $a_t^\xi$  term would easily overwhelm the other term.

Let us conclude with the following comment. Kohlhas and Walther (2019) have used a model of the type described above to show that asymmetric attention allocation to various components of the outcome may help reconcile the form of belief over-reaction documented in their paper with the form of belief under-reaction documented in CG. Our network extension in Section VIII allows one to consider a multi-sector economy in which different sectors have different exposures to the aggregate shock, either directly or indirectly via differential GE effects. This may provide a more detailed micro-foundation for pro- and counter-cyclical components of economic activity, along the lines suggested by the aforementioned paper. And it could help study the role of asymmetries in GE feedbacks, similarly in spirit to what we do in our HANK application in Section VII.<sup>17</sup>

## J. ADDITIONAL PROOFS

### *Proof of Lemma D.1, Lemma D.2, and Proposition 8*

The demand schedule faced by an individual firm  $i$  in market  $m$  is given by

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

where  $\psi$  and  $\varepsilon$  are within- and across-market elasticities of substitution, respectively. The price index in market  $m$  and the aggregate price index are defined as

$$P_{m,t} = \left( \frac{1}{N} \sum_j P_{j,m,t}^{1-\psi} \right)^{\frac{1}{1-\psi}}, \quad P_t = \left[ \int_m P_{m,t}^{1-\varepsilon} \right]^{\frac{1}{1-\varepsilon}}.$$

In the absence of nominal rigidity and informational frictions, an individual firm  $i$  in market  $m$  sets its price to maximize its profit in the current period

$$\max_{P_{i,m,t}} P_{i,m,t} Y_{i,m,t} - P_t C(Y_t) Y_{i,m,t},$$

where  $C(Y_t)$  is the marginal real cost which depends on the aggregate economic condition. Using the following properties

$$\frac{\partial Y_{i,m,t}}{\partial P_{i,m,t}} = -\psi \frac{Y_{i,m,t}}{P_{i,m,t}} + (\psi - \varepsilon) \frac{Y_{i,m,t}}{P_{m,t}} \frac{\partial P_{m,t}}{\partial P_{i,m,t}}, \quad \text{and} \quad \frac{\partial P_{m,t}}{\partial P_{i,m,t}} = \frac{1}{N} P_{m,t}^\psi P_{i,m,t}^{-\psi},$$

<sup>17</sup>There, asymmetric GE feedbacks emerge because of heterogeneous MPCs and heterogeneous exposures of income to business-cycle fluctuations.

the first-order condition is

$$(1 - \psi) \frac{P_{i,m,t}}{P_t} + \psi C(Y_t) = \frac{\varepsilon - \psi}{N} \left( \frac{P_{i,m,t}}{P_t} - C(Y_t) \right) \left( \frac{P_{i,m,t}}{P_{m,t}} \right)^{1-\psi} = 0.$$

We assume that  $C(Y_t) = C \exp(\text{mc}_t)$  where  $\text{mc}_t$  follows an AR(1) process

$$\text{mc}_t = \rho \text{mc}_{t-1} + \eta_t.$$

In steady state where  $\text{mc}_t = 0$  and  $P_{i,m,t} = P_{m,t} = P_t$ , it follows that

$$C = \frac{\psi - 1 + \frac{\varepsilon - \psi}{N}}{\psi + \frac{\varepsilon - \psi}{N}}.$$

The log-linearized version of the first-order condition is

$$(1 - \psi)(p_{i,m,t} - p_t) + \psi C \text{mc}_t = \frac{\varepsilon - \psi}{N} \left( p_{i,m,t} - p_t - C \text{mc}_t + (1 - C)(1 - \psi)(p_{i,m,t} - p_{m,t}) \right),$$

which leads to the following best response

$$p_{i,m,t} = \varphi \text{mc}_t + \alpha_N p_{m,t} + (1 - \alpha_N) p_t,$$

where  $\alpha_N$  is given by

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

Turn to the environment where there is nominal rigidity and incomplete information. The problem of a firm that can reset its price becomes

$$\max_{P_{i,m,t}} \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,m,t} \left[ P_{i,m,t} Y_{i,m,t+k} - P_{t+k} C(Y_{t+k}) Y_{i,m,t+k} \right],$$

and the linearized first-order condition becomes

$$p_{i,m,t}^* = (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \varphi \mathbb{E}_{i,m,t} [\text{mc}_{t+k}] + (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{i,m,t} [\alpha_N p_{m,t+k} + (1 - \alpha_N) p_{t+k}].$$

Under the assumption that all firms share the same information within the market, all newly set prices within a market are identical. Denote the newly set price in market  $m$  as  $p_{m,t}^*$ , and it satisfies

$$p_{m,t}^* = (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \varphi \mathbb{E}_{m,t} [\text{mc}_{t+k}] + (1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{m,t} [\alpha_N p_{m,t+k} + (1 - \alpha_N) p_{t+k}].$$

Denote  $\pi_{m,t}$  as the inflation rate in market  $m$ . Subtracting  $p_{m,t-1}$  from both sides of the equation above leads to

$$\begin{aligned} \pi_{m,t} = & (1 - \theta)(1 - \chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \varphi \mathbb{E}_{m,t} [\text{mc}_{t+k}] + \alpha_N (1 - \theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{m,t} [\pi_{m,t+k}] \\ & + (1 - \theta)(1 - \alpha_N) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{m,t} [\pi_{t+k}] + (1 - \theta)(1 - \alpha_N)(p_{t-1} - p_{m,t-1}). \end{aligned}$$

To proceed, consider the following alternative inflation definition in market  $m$

$$\begin{aligned}\tilde{\pi}_{m,t} = & (1-\theta)(1-\chi\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \varphi \mathbb{E}_{m,t}[\text{mc}_{t+k}] + \alpha_N(1-\theta) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{m,t}[\tilde{\pi}_{m,t+k}] \\ & + (1-\theta)(1-\alpha_N) \sum_{k=0}^{\infty} (\chi\theta)^k \mathbb{E}_{m,t}[\pi_{t+k}].\end{aligned}$$

Since the aggregate inflation under these two models are identical ( $\int_m \pi_{m,t} = \int_m \tilde{\pi}_{m,t}$ ), we can derive the aggregate inflation dynamics from the latter. By the law of iterated expectations, we have

$$\begin{aligned}\tilde{\pi}_{m,t} = & \mathbb{E}_{m,t} \left[ \frac{(1-\theta)(1-\chi\theta)\varphi \text{mc}_t + (1-\theta)(1-\alpha_N)\pi_t}{1-\chi\theta L^{-1}} \right] + (1-\theta)\alpha_N \mathbb{E}_{m,t} \left[ \frac{\tilde{\pi}_{m,t}}{1-\chi\theta L^{-1}} \right] \\ = & \mathbb{E}_{m,t} \left[ \frac{(1-\theta)(1-\chi\theta)\varphi \text{mc}_t + (1-\theta)(1-\alpha_N)\pi_t}{1-\chi\theta L^{-1}} \left( 1 - \frac{(1-\theta)\alpha_N}{1-\chi\theta L^{-1}} \right)^{-1} \right] \\ = & \frac{1}{1-(1-\theta)\alpha_N} \sum_{k=0}^{\infty} \left( \frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \mathbb{E}_{m,t}[(1-\theta)(1-\chi\theta)\varphi \text{mc}_{t+k} + (1-\theta)(1-\alpha_N)\pi_{t+k}].\end{aligned}$$

Aggregating across markets and using the assumption that firms can observe current inflation, it follows that the aggregate inflation satisfies

$$\pi_t = \kappa \sum_{k=0}^{\infty} \left( \frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \bar{\mathbb{E}}_t[\text{mc}_{t+k}] + \frac{\chi(1-\theta)(1-\alpha_N)}{1-(1-\theta)\alpha_N} \sum_{k=0}^{\infty} \left( \frac{\chi\theta}{1-(1-\theta)\alpha_N} \right)^k \bar{\mathbb{E}}_t[\pi_{t+k+1}],$$

where  $\kappa = \frac{(1-\chi\theta)(1-\theta)\varphi}{\theta}$ . Mapping the fixed point problem above to our baseline framework, the aggregate outcome is the result of the following forward-looking game

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

where

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

with  $\beta + \gamma = \chi$ . Note that  $\gamma$  is decreasing in  $\alpha_N$ . To show that  $\gamma$  is increasing in  $N$ , it is sufficient to show that  $\alpha_N$  is decreasing in  $N$ . When  $\psi > \varepsilon > 1$ , and  $N \geq 2$

$$\begin{aligned}\frac{\partial \alpha_N}{\partial N} &= \frac{(\psi-1)(\psi-\varepsilon)(\psi^2 + \varepsilon^2 - 2\psi\varepsilon - \psi N^2(\psi-1))}{(\psi(N^2(\psi-1) - (N-1)\psi) + (N-2)\psi\varepsilon + \varepsilon^2)^2} \\ &< \frac{(\psi-1)(\psi-\varepsilon)(\psi^2 - \psi + \varepsilon^2 - \varepsilon - N^2(\psi^2 - \psi))}{(\psi(N^2(\psi-1) - (N-1)\psi) + (N-2)\psi\varepsilon + \varepsilon^2)^2} \\ &< 0,\end{aligned}$$

which completes the proof.



*Proof of Proposition 12*

The signal process can be represented as

$$\begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} \tau_\varepsilon^{-1/2} & 0 & \frac{1}{1-\rho L} \\ 0 & \tau_u^{-1/2} & \frac{1}{1-\rho L} \end{bmatrix}}_{\equiv \mathbf{M}(L)} \underbrace{\begin{bmatrix} \widehat{\varepsilon}_t \\ \widehat{u}_{i,t} \\ \widehat{\eta}_t \end{bmatrix}}_{\equiv \widehat{\mathbf{s}}_{i,t}}.$$

where  $\widehat{\mathbf{s}}_{i,t}$  is a vector of standardized normal random variables. The auto-covariance generating function for the signal process is

$$\mathbf{M}(L) \mathbf{M}'(L^{-1}) = \frac{1}{(L-\rho)(1-\rho L)} \begin{bmatrix} L + \frac{(L-\rho)(1-\rho L)}{\tau_\varepsilon} & L \\ L & L + \frac{(L-\rho)(1-\rho L)}{\tau_u} \end{bmatrix}.$$

In order to apply the Wiener-Hopf prediction formula we need to obtain the canonical factorization. Let  $\lambda$  be the inside root of the determinant of  $\mathbf{M}(L) \mathbf{M}'(L^{-1})$

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

Then the fundamental representation is given by

$$\begin{aligned} \mathbf{B}(z)^{-1} &= \frac{1}{1-\lambda z} \begin{bmatrix} 1 - \frac{\tau_\varepsilon \rho + \lambda \tau_u}{\tau_\varepsilon + \tau_u} z & \frac{\tau_u(\lambda - \rho)}{\tau_\varepsilon + \tau_u} z \\ \frac{\tau_\varepsilon(\lambda - \rho)}{\tau_\varepsilon + \tau_u} z & 1 - \frac{\tau_u \rho + \lambda \tau_\varepsilon}{\tau_\varepsilon + \tau_u} z \end{bmatrix}, \\ \mathbf{V}^{-1} &= \frac{\tau_\varepsilon \tau_u}{\rho(\tau_\varepsilon + \tau_u)} \begin{bmatrix} \frac{\tau_u \rho + \lambda \tau_\varepsilon}{\tau_u} & \lambda - \rho \\ \lambda - \rho & \frac{\tau_\varepsilon \rho + \lambda \tau_u}{\tau_\varepsilon} \end{bmatrix}, \end{aligned}$$

which satisfies

$$\mathbf{B}(L) \mathbf{V} \mathbf{B}'(L^{-1}) = \mathbf{M}(L) \mathbf{M}'(L^{-1}).$$

Applying the Wiener-Hopf prediction formula, the forecast of  $\xi_t$  is given by

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

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  $a_t = g(L)\xi_t + h_1(L)\varepsilon_t$ . The forecast about  $a_{t+1}$  is given by

$$\begin{aligned} \mathbb{E}_{i,t}[a_{t+1}] &= \left[ \begin{bmatrix} \tau_\varepsilon^{-1/2} L^{-1} h_1(L) & 0 & \frac{L^{-1} g(L)}{1-\rho L} \end{bmatrix} \mathbf{M}'(L^{-1}) \mathbf{B}'(L^{-1})^{-1} \right]_+ \mathbf{V}^{-1} \mathbf{B}(L)^{-1} \begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix} \\ &= \left\{ \frac{[(\rho\tau_u + \lambda\tau_\varepsilon + \lambda\rho(\lambda\tau_u + \rho\tau_\varepsilon))L - \lambda\rho(\tau_u + \tau_\varepsilon)(1+L^2)]h_1(L) - \tau_u(\lambda-\rho)(1-\rho\lambda)Lh_1(L)}{\rho(\tau_u + \tau_\varepsilon)L(L-\lambda)(1-\lambda L)} \right\} \end{aligned}$$

$$\begin{aligned}
& - \frac{[\tau_\varepsilon (\rho - \lambda) (1 - \rho L) L h_1 (\lambda) \quad \tau_u (\rho - \lambda) (1 - \rho L) L h_1 (\lambda)]}{\rho (\tau_u + \tau_\varepsilon) L (L - \lambda) (1 - \lambda L)} \\
& - \frac{[\rho (L - \lambda) ((\lambda \tau_u + \rho \tau_\varepsilon) L - (\tau_u + \tau_\varepsilon)) h_1 (0) \quad \tau_u (\rho - \lambda) L \rho (L - \lambda) h_1 (0)]}{\rho (\tau_u + \tau_\varepsilon) L (L - \lambda) (1 - \lambda L)} \\
& + \frac{\lambda ((1 - \rho \lambda) g (L) - (1 - \rho L) g (\lambda)) [\tau_\varepsilon \quad \tau_u]'}{\rho (1 - \rho \lambda) (L - \lambda) (1 - \lambda L)} \left. \vphantom{\frac{\lambda ((1 - \rho \lambda) g (L) - (1 - \rho L) g (\lambda)) [\tau_\varepsilon \quad \tau_u]'}{\rho (1 - \rho \lambda) (L - \lambda) (1 - \lambda L)}}} \right\} \begin{bmatrix} z_t \\ x_{i,t} \end{bmatrix}.
\end{aligned}$$

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

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

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

$$\mathbf{A} (L) \begin{bmatrix} h_1 (L) \\ h_2 (L) \end{bmatrix} = \mathbf{d} (L),$$

where<sup>18</sup>

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

and

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

<sup>18</sup>We have used the following identities to simplify the expressions

$$\begin{aligned}
\rho \tau_u + \lambda \tau_\varepsilon + \lambda \rho (\lambda \tau_u + \rho \tau_\varepsilon) + \lambda \tau_\varepsilon (\tau_u + \tau_\varepsilon) &= \rho (1 + \lambda^2) (\tau_u + \tau_\varepsilon), \\
\rho \tau_\varepsilon + \lambda \tau_u + \lambda \rho (\lambda \tau_\varepsilon + \rho \tau_u) + \lambda \tau_u (\tau_u + \tau_\varepsilon) &= \rho (1 + \lambda^2) (\tau_u + \tau_\varepsilon).
\end{aligned}$$

with

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

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  $\omega_1$  to  $\omega_4$ , with  $|\omega_4| > 1$  and the others being less than 1 in absolute value. We choose  $\varphi_1$ ,  $\varphi_2$ , and  $\varphi_3$  to remove the inside poles of  $h_1(L)$  at  $\omega_1$  to  $\omega_3$ . 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  $\sigma^{-2} = \sigma_u^{-2} + \sigma_\varepsilon^{-2}$ . The aggregate outcome,  $a_t = (h_1(L) + h_2(L))\xi_t + h_1(L)\varepsilon_t$ , is

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

In terms of comparative statics, note that

$$\frac{\partial C(\vartheta^{-1})}{\partial \chi} = \frac{\chi}{\rho\sigma^2} > 0.$$

By the same logic in the proof of Proposition 5, it follows that  $\vartheta$  is decreasing in  $\chi$ .

*Proof of Proposition 13*

This follows directly from the analysis in the main text.

*Proof of Proposition 14*

First, let us prove  $g_k < \widehat{g}_k$ . Recall that  $\{g_k\}$  is given by

$$g_k = \sum_{h=0}^{\infty} \gamma^h \lambda_k \lambda_{k+1} \dots \lambda_{k+h} \rho_{k+h}.$$

Clearly,

$$0 < g_k < \sum_{h=0}^{\infty} \gamma^h \lambda_k \rho_{k+h} = \widehat{g}_k,$$

which proves the first property. If  $\lim_{k \rightarrow \infty} \lambda_k = 1$  and  $\sum_{h=0}^{\infty} \gamma^h \rho_{k+h}$  exists for all  $k$ , then it follows that

$$\lim_{k \rightarrow \infty} \frac{\widehat{g}_k}{g_k} = \frac{\lim_{k \rightarrow \infty} \sum_{h=0}^{\infty} \gamma^h \rho_{k+h}}{\lim_{k \rightarrow \infty} \sum_{h=0}^{\infty} \gamma^h \rho_{k+h}} = 1.$$

Next, let us prove that  $\frac{g_{k+1}}{g_k} > \frac{\widehat{g}_{k+1}}{\widehat{g}_k}$ . By definition,

$$\begin{aligned} \frac{\widehat{g}_{k+1}}{\widehat{g}_k} &= \frac{\lambda_{k+1} \sum_{h=0}^{\infty} \gamma^h \rho_{k+h+1}}{\lambda_k \sum_{h=0}^{\infty} \gamma^h \rho_{k+h}}, \\ \frac{g_{k+1}}{g_k} &= \frac{\lambda_{k+1} \sum_{h=0}^{\infty} \gamma^h \lambda_{k+2} \dots \lambda_{k+h+1} \rho_{k+h+1}}{\lambda_k \sum_{h=0}^{\infty} \gamma^h \lambda_{k+1} \dots \lambda_{k+h} \rho_{k+h}}. \end{aligned}$$

Since  $\{\lambda_k\}$  is strictly increasing and  $\rho_k > 0$ , we have

$$\frac{g_{k+1}}{g_k} \Big/ \frac{\widehat{g}_{k+1}}{\widehat{g}_k} > \frac{\sum_{h=0}^{\infty} \gamma^h \lambda_{k+1} \dots \lambda_{k+h} \rho_{k+h+1}}{\sum_{h=0}^{\infty} \gamma^h \lambda_{k+1} \dots \lambda_{k+h} \rho_{k+h}} \Big/ \frac{\sum_{h=0}^{\infty} \gamma^h \rho_{k+h+1}}{\sum_{h=0}^{\infty} \gamma^h \rho_{k+h}}.$$

It is sufficient to show that the term on the right-hand side is greater than 1. To proceed, we start with the following observation. If  $\theta_1 \geq \theta_2 > 0$ , and  $\frac{y_2}{y_1+y_2} \geq \frac{x_2}{x_1+x_2}$ , then

$$(J1) \quad \frac{x_1 \theta_1 + x_2 \theta_2}{x_1 + x_2} \geq \frac{y_1 \theta_1 + y_2 \theta_2}{y_1 + y_2}.$$

Note that

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

and

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

In what follows, we will show by induction that

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

We first establish the following inequality

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

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  $y_2 = \gamma \frac{\rho_{k+1}}{\rho_k},$  and applying inequality (J1). By assumption,  $\frac{\rho_k \rho_{k+2}}{\rho_{k+1}^2} \leq 1,$  which implies  $\theta_1 \geq \theta_2 > 0.$  Meanwhile,

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

Now suppose that

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

and we need to show

$$(J2) \quad \frac{1 + \gamma \lambda_{k+1} \frac{\rho_{k+2}}{\rho_{k+1}} + \dots + \gamma^{n-1} \lambda_{k+1} \dots \lambda_{k+n-1} \frac{\rho_{k+n}}{\rho_{k+1}} + \gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n+1}}{\rho_{k+1}}}{1 + \gamma \lambda_{k+1} \frac{\rho_{k+1}}{\rho_k} + \dots + \gamma^{n-1} \lambda_{k+1} \dots \lambda_{k+n-1} \frac{\rho_{k+n-1}}{\rho_k} + \gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n}}{\rho_k}} \\ \geq \frac{1 + \gamma \frac{\rho_{k+2}}{\rho_{k+1}} + \dots + \gamma^{n-1} \frac{\rho_{k+n}}{\rho_{k+1}} + \gamma^n \frac{\rho_{k+n+1}}{\rho_{k+1}}}{1 + \gamma \frac{\rho_{k+1}}{\rho_k} + \dots + \gamma^{n-1} \frac{\rho_{k+n-1}}{\rho_k} + \gamma^n \frac{\rho_{k+n}}{\rho_k}}.$$

Again, to apply (J1), 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

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

and

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

To establish (J2), 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

$$\frac{\theta_1}{\theta_2} = \frac{1 + \gamma \frac{\rho_{k+1}}{\rho_k} \frac{\rho_{k+2} \rho_k}{\rho_{k+1}^2} + \dots + \gamma^{n-1} \frac{\rho_{k+n-1}}{\rho_k} \frac{\rho_{k+n} \rho_k}{\rho_{k+1} \rho_{k+n-1}}}{\theta_2 + \gamma \frac{\rho_{k+1}}{\rho_k} \theta_2 + \dots + \gamma^{n-1} \frac{\rho_{k+n-1}}{\rho_k} \theta_2}.$$

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

$$\frac{x_2}{x_1 + x_2} \\ = \frac{\gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n}}{\rho_k}}{1 + \gamma \lambda_{k+1} \frac{\rho_{k+1}}{\rho_k} + \dots + \gamma^{n-1} \lambda_{k+1} \dots \lambda_{k+n-1} \frac{\rho_{k+n-1}}{\rho_k} + \gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n}}{\rho_k}} \\ \leq \frac{\gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n}}{\rho_k}}{\lambda_{k+1} \dots \lambda_{k+n} + \gamma \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+1}}{\rho_k} + \dots + \gamma^{n-1} \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n-1}}{\rho_k} + \gamma^n \lambda_{k+1} \dots \lambda_{k+n} \frac{\rho_{k+n}}{\rho_k}} \\ = \frac{y_2}{y_1 + y_2}.$$

This completes the proof that  $\frac{g_{k+1}}{g_k} > \frac{\hat{g}_{k+1}}{\hat{g}_k}$ .

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.** 2018. “Forward Guidance without Common Knowledge.” *American Economic Review*, 108(9): 2477–2512.
- Angeletos, George-Marios, Fabrice Collard, and Harris Dellas.** 2020. “Business-Cycle Anatomy.” *American Economic Review*, 110(10): 3030–70.
- 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.
- 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.
- 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.** 2015. “Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts.” *American Economic Review*, 105(8): 2644–78.
- 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.
- Golosov, Mikhail, and Robert E Lucas Jr.** 2007. “Menu Costs and Phillips Curves.” *Journal of Political Economy*, 115(2): 171–199.
- 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.
- Jung, Jeeman, and Robert J. Shiller.** 2005. “Samuelson’s Dictum and the Stock Market.” *Economic Inquiry*, 43(2): 221–228.
- Kasa, Kenneth, Todd B. Walker, and Charles H. Whiteman.** 2014. “Heterogeneous Beliefs and Tests of Present Value Models.” *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.
- Kohlhas, Alexandre, and Ansgar Walther.** 2019. “Asymmetric Attention.” *IIES miméo*.
- Lucas, Robert E. Jr.** 1972. “Expectations and the Neutrality of Money.” *Journal of Economic Theory*, 4(2): 103–124.
- Lucas, Robert E. Jr.** 1976. “Econometric Policy Evaluation: A Critique.” *Carnegie-Rochester Conference Series on Public Policy*, 1: 19 – 46.
- Maćkowiak, Bartosz, and Mirko Wiederholt.** 2009. “Optimal Sticky Prices under Rational Inattention.” *American Economic Review*, 99(3): 769–803.
- 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.
- Melosi, Leonardo.** 2016. “Signalling effects of monetary policy.” *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.** 2002. “Social Value of Public Information.” *American Economic Review*, 92(5): 1521–1534.
- Morris, Stephen, and Hyun Song Shin.** 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.
- Nimark, Kristoffer.** 2017. “Dynamic Higher Order Expectations.” *Cornell Univeristy mimeo*.
- Romer, Christina D, and David H Romer.** 2004. “A new measure of monetary shocks: Derivation and implications.” *American Economic Review*, 94(4): 1055–1084.
- Singleton, Kenneth J.** 1987. “Asset Prices in a Time-series Model with Disparately Informed, Competitive Traders.” In *New Approaches to Monetary Economics*, ed. William A. Barnett and Kenneth J. Singleton, 249–272. Cambridge University Press.
- Smets, Frank, and Rafael Wouters.** 2007. “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” *American Economic Review*, 97(3): 586–606.
- Vives, Xavier, and Liyan Yang.** 2017. “Costly Interpretation of Asset Prices.” IESE/University of Toronto mimeo.
- 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.” University of Munich miméo.