A Hierarchical Representation of Dispersed Information*

Giacomo Rondina[†] Todd B. Walker[‡]

August 2026

Abstract

We establish an aggregate equivalence between dispersed and hierarchical-information economies in canonical beauty-contest models. We refine the hierarchy with a "Russian Doll" construction that permits exact aggregate equivalence while making individual actions arbitrarily close to those in the dispersed economy. The equivalence is useful for three reasons. First, it turns higher-order beliefs into tractable objects, available in closed form. Second, it delivers identification of the information structure itself from cross-sectional moments, an object that does not exist when information is symmetrically dispersed. Third, it extends the equivalence to endogenous signal extraction, permitting an analytical characterization of information transmission.

keywords: higher-order beliefs, dispersed information, identification
JEL: D83, D84, E31, E32, E71

*This paper extends and subsumes "Dispersed and Hierarchical Economics with Dynamic Signal Extraction: An Equivalence Result" by the same authors. We thank Manuel Amador, Marios Angeletos, Ryan Chahrour, Zhen Huo, Jennifer La'O, Eric Leeper, Kristoffer Nimark, Tom Sargent, Martin Schneider, Venky Venkateswaran, Pierre-Olivier Weill, Mirko Wiederholt, Chuck Whiteman and Dalton Zhang for helpful comments on an earlier draft.

[†]Department of Economics, UCSD, grondina@ucsd.edu.
[‡]Department of Economics, Indiana University, walkertb@iu.edu

1 Introduction

Dispersed-information models require agents to forecast not only fundamentals but also the forecasts of others. This higher-order-belief channel is central to many explanations of sluggish aggregate adjustment, persistent forecast disagreement, and systematic departures from full-information rational expectations (Angeletos and Lian, 2016). Recent work has produced sharp characterizations of aggregate dynamics in canonical environments, often in closed form (Huo and Pedroni, 2020; Angeletos and Huo, 2021; Huo and Takayama, 2025). However, the cross-sectional objects that dispersed information is meant to explain—the distribution of individual beliefs and actions, higher-order belief dynamics, and moments of forecast disagreement—remain difficult to compute because each depends on an infinite hierarchy of average expectations whose ARMA order grows with belief order (Angeletos and Huo, 2021; Huo and Takayama, 2025), and which need not admit any finite representation once information is endogenous (Chahrour and Jurado, 2025). Second, and more fundamentally, aggregate data do not identify how information is distributed. Economies with very different cross-sectional distributions of information can generate identical aggregate dynamics, identical average forecasts, and identical values of standard empirical measures of information rigidity (as we document below).

This paper addresses both problems by constructing hierarchical representations of dispersed information economies. We operate within the canonical beauty contest model of Morris and Shin (2002) and Angeletos and Huo (2021). We first show that any dispersed-information economy in this class is exactly aggregate-equivalent to one in which information is nested (i.e. agents can be ranked by information quality). This nesting delivers tractability in that it yields closed-form expressions for the entire hierarchy of higher-order beliefs. We then derive refinements, which we call "Russian Doll" hierarchies. Each refinement preserves exact aggregate equivalence, while the type-specific individual actions converge to those in the original dispersed-information economy. The end result is an equilibrium that is aggregate-equivalent to the dispersed economy with individual actions indistinguishable from it.

Our sufficient-statistic property is also the source of an identification challenge. Because aggregates depend only on the mean of the information distribution, the equivalence class is large; a continuum of nested structures, of any depth, with any admissible shares and loadings, generates the same aggregate action, the same average forecast at every horizon, the same population-average belief hierarchy, and the same behavioral interpretation. Aggregate data therefore identify the common forecasting filter but not the *distribution* of information. Moreover, this identification challenge cannot be addressed when heterogeneous information is symmetrically distributed as this yields a degenerate distribution of information.

However once information is hierarchical, the distribution becomes non-degenerate and its cross-sectional moments can recover the number of types, population shares, and information loadings. This identification survives the Russian Doll approximation because for every positive informational separation, however small, the finite hierarchy remains identified even though its individual conditional distributions can be arbitrarily close to the dispersed-information benchmark. Theorem 3 offers global identification in this framework.

Our third contribution applies these ideas to models with endogenous information acquisition. Learning from equilibrium prices and actions is central to this literature (Lucas, 1972; Townsend, 1983; Grossman and Stiglitz, 1980; Hellwig, 1980), and yet finite-state representations need not survive endogenous information aggregation (Huo and Takayama, 2025; Chahrour and Jurado, 2025). These challenges severely limit the applicability of promising models.

We allow for endogenous learning in the dynamic beauty contest model and derive restrictions on the hierarchical information structure that delivers closed form expressions for all equilibrium objects. Our Russian Doll construction then preserves this aggregate equilibrium while making each type's conditional action distribution arbitrarily close to a symmetrically dispersed-information benchmark. We then study information transmission in this framework and derive the exact full-revelation threshold and show that informed agents transmit additional private information by acting on their anticipation of others' forecast errors. In our numerical illustration, suppressing this higher-order-belief channel nearly doubles the informed share required for full revelation.

1.1 RELATION TO THE LITERATURE

We work within the canonical beauty-contest model of Morris and Shin (2002) and Angeletos and Huo (2021). The dynamic version nests the dynamic IS curve, the New Keynesian Phillips curve, standard asset-pricing Euler equations, and reduced forms commonly used in heterogeneous-agent and production-network models. Establishing an aggregate equivalence in this benchmark environment therefore has implications for a broader range of applications than what is studied here. Moreover, by extending our results to endogenous information, we open the equivalence to a much larger class of environments in which agents learn from equilibrium outcomes (Lucas, 1972; Grossman and Stiglitz, 1980; Hellwig, 1980; Angeletos and Werning, 2006; Amador and Weill, 2010). Our modeling techniques can be applied to a broader literature on dispersed information and imperfect common knowledge, in which incomplete information generates sluggish adjustment, higher-order uncertainty, and forecast disagreement (Woodford, 2003; Lorenzoni, 2009; Angeletos and La'O, 2013; Melosi, 2014; Nimark, 2017); Angeletos and Lian (2016) provide an overview. Related informational-rigidity models include Mankiw and Reis (2002) and Mackowiak and Wiederholt (2009).

Many recent papers derive aggregate-equivalence results to simplify and better understand beauty contest models. Huo and Pedroni (2020) characterize the equilibrium policy rule through a single-agent forecasting problem with modified idiosyncratic precision; Angeletos and Huo (2021) represent the aggregate effects of dispersed information as myopia and anchoring; and Huo and Takayama (2025) show that, when signals are finite ARMA, the weighted sum of higher-order expectations that enters the aggregate equilibrium admits a finite-state representation. These papers reveal that aggregate outcomes can be much simpler than the underlying hierarchy of beliefs. We demonstrate that such reductions are also statements about identification. If aggregates depend on the information structure only through a low-dimensional summary, then this implies aggregate data cannot identify information flows. Our nested information structure makes the hierarchy itself tractable and provides a cross-sectional distribution that achieves global identification.

Finally, the paper speaks to the empirical literature on expectations.[1] Coibion and Gorodnichenko

[1] Our paper also contributes to the moving-average / frequency-domain approach to information frictions (Futia, 1981; Taub, 1989; Kasa, 2000; Walker, 2007; Rondina, 2008; Acharya, 2013; Kasa et al., 2014; Tan and Walker, 2015; Rondina and Walker,

(2015) introduce the forecast-error–revision coefficient that has become the standard aggregate diagnostic of information rigidity. Angeletos and Huo (2021) interpret this friction through a behavioral lens of myopia and anchoring. Our equivalence shows that these statistics do not identify how information is distributed as every admissible hierarchy preserves the Coibion–Gorodnichenko coefficient, and the myopia and anchoring coefficients. The mapping from information distributions to aggregate evidence is many-to-one. Cross-sectional data is needed for identification. We show that state-dependent disagreement identifies information inequality, while higher moments of individual behavior globally recover the shares and ordered information intensities of any finite hierarchy.

2 MODEL SETUP

We work within the canonical beauty contest metaphor of Keynes, formalized in Morris and Shin (2002) and Angeletos and Huo (2021). These models offer a tractable, micro-founded approach to strategic complementarities and higher-order beliefs. Angeletos and Huo (2021) extend the static Morris-Shin structure, nesting it inside a dynamic representative-agent setup that maps directly onto the New Keynesian Phillips curve, the dynamic IS equation, asset-pricing relations, heterogeneous-agent (HANK) economies, and production networks. They show how this framework can offer insights on the empirical sluggishness of macro data, mappings from survey expectations to macro dynamics, and the nesting of DSGE, HANK, and network literatures. Hence, our results pertain to a wide variety of settings.

2.1 STATIC MODEL Consider the static Beauty Contest Model of Morris and Shin (2002) (MS, hereafter) where a continuum of agents, indexed by i, chooses action a_i according to the optimal linear response rule, $a_i = (1-r)\mathbb{E}_i[\theta] + r\mathbb{E}_i[\bar{a}]$, with $\bar{a} \equiv \int_0^1 a_j\, dj$ and where θ is an unknown fundamental. The parameter $r \in (0,1)$ governs the strength of strategic complementarities, where $(1-r)\mathbb{E}_i[\theta]$ captures the accuracy motive (keeping a_i close to θ), while $r\mathbb{E}_i[\bar{a}]$ captures the beauty-contest motive (keeping a_i close to the average action of others). Each agent observes a public signal $y = \theta + \eta$ of the fundamental θ with precision α (i.e., $\text{Var}(\eta) = 1/\alpha$) and a private signal $x_i = \theta + \varepsilon_i$ of precision β (i.e., $\text{Var}(\varepsilon_i) = 1/\beta$), where η and all ε_i are independent and normally distributed. The posterior mean of the fundamental is the well-known weighted average

$$\mathbb{E}_i[\theta \mid y, x_i] = \left(\frac{\alpha}{\alpha+\beta}\right) y + \left(\frac{\beta}{\alpha+\beta}\right) x_i \tag{1}$$

MS use a two-step procedure to prove the existence and uniqueness of the equilibrium. First, they guess that all agents follow the linear strategy $a_j = Kx_j + (1-K)y$, and then show $K = (\beta(1-r))/(\alpha+\beta(1-r))$. The linear guess is verified with equilibrium action

$$a_i = \left(\frac{\beta(1-r)}{\alpha+\beta(1-r)}\right) x_i + \left(\frac{\alpha}{\alpha+\beta(1-r)}\right) y \tag{2}$$

A "brute force" induction is then employed to prove consistency of this equilibrium with the higher-order beliefs equilibrium, making it unique. MS show that recursive substitution of the averaged best-response

2021; Acharya et al., 2021; Miao et al., 2021; Han et al., 2022; Huo and Takayama, 2025; Jurado, 2023; Chahrour and Jurado, 2025).

function $\bar{a} = (1-r)\overline{\mathbb{E}}[\theta] + r\overline{\mathbb{E}}[\bar{a}]$ yields

$$\bar{a} = (1-r)\sum_{k=1}^{\infty} r^{k-1}\overline{\mathbb{E}}^{k}[\theta] + \lim_{n\to\infty} r^{n+1}\overline{\mathbb{E}}[\bar{a}] = (1-r)\sum_{k=1}^{\infty} r^{k-1}\,\overline{\mathbb{E}}^{k}[\theta] \tag{3}$$

where the last equality holds given $r \in (0,1)$ and standard transversality conditions, and where $\overline{\mathbb{E}}^{k}[\theta]$ is the k-th-order average expectation ($\overline{\mathbb{E}}^{0}[\theta] \equiv \theta$). Using an induction on k, Lemma 1 of MS demonstrates that for any k,

$$\overline{\mathbb{E}}^{k}[\theta] = (1-\mu^{k})y + \mu^{k}\theta \qquad \mathbb{E}_i\left(\overline{\mathbb{E}}^{k}[\theta]\right) = (1-\mu^{k+1})y + \mu^{k+1}x_i \tag{4}$$

where $\mu \equiv \beta/(\alpha+\beta)$. Substituting (4) into (3) yields (2) and uniqueness follows.

Dispersed-Hierarchical Aggregate Equivalence. The dispersed-information economy of MS has an *aggregate* equivalent where information is hierarchical or nested (i.e., uninformed agents' information is a strict subset of the informed). The average expectation integrates over (1)

$$\overline{\mathbb{E}}[\theta] = \int_0^1 \mathbb{E}_i[\theta]\,di = (1-\mu)\,y + \mu\int_0^1 x_i\,di = (1-\mu)\,y + \mu(\theta + \int_0^1 \varepsilon_i\,di) = \mu\theta + (1-\mu)\,y \tag{5}$$

where the last equality follows from the law of large numbers $\int_0^1 \varepsilon_i\,di = 0$. Equation (5) admits a natural signal-extraction interpretation. At the individual level, the expectation $\mathbb{E}_i[\theta \mid y, x_i] = (1-\mu)\,y + \mu\,x_i$ can heuristically be read as if the agent places weight $1-\mu = \alpha/(\alpha+\beta)$ on the public signal and weight $\mu = \beta/(\alpha+\beta)$ on the private signal. We operationalize this interpretation to define a hierarchical-information economy in which a fraction μ of *informed* agents observe $\Omega^{I} = \{y,\,\theta\}$ and the remaining fraction $1-\mu$ of *uninformed* agents observe only the public signal $\Omega^{U} = \{y\}$. In this hierarchical economy, the average expectation is $\overline{\mathbb{E}}[\theta] = \mu\theta + (1-\mu)\,y$, which coincides exactly with (5) when $\mu = \beta/(\alpha+\beta)$.

Moreover, because every uninformed agent's information set is strictly nested within that of the informed, the informed agents can perfectly anticipate the uninformed's expectation of θ. The informed agents forecast the forecast *errors* of the uninformed. Higher-order beliefs become simple calculations

$$\mathbb{E}^{I}\left[\overline{\mathbb{E}}^{k}[\theta]\right] = y + \mu^{k}(\theta - y) \qquad \mathbb{E}^{U}\left[\overline{\mathbb{E}}^{k}[\theta]\right] = y$$

for all $k \geq 0$ and where $\mathbb{E}^{I}$ and $\mathbb{E}^{U}$ are the conditional expectations of the informed and uninformed. Averaging across types then yields the analytical expression (4). This hierarchy exactly matches the higher-order-belief structure of MS (Equation (3)). Hence we have shown the following:

Proposition 1. *The dispersed-information economy of Morris and Shin (2002) is aggregate-equivalent to an economy in which a fraction* μ *of agents observe* (y,θ) *and a fraction* $1-\mu$ *observe only* (y) *when* $\mu \equiv \beta/(\alpha+\beta)$. *In both economies,* $\overline{\mathbb{E}}^{k}[\theta] = (1-\mu^{k})y + \mu^{k}\theta$ *for* $k \geq 1$ *and therefore the aggregate action is identical and given by*

$$\bar{a} = y + \frac{\mu(1-r)}{1-r\mu}(\theta - y) = \left(\frac{\beta(1-r)}{\alpha+\beta(1-r)}\right)\theta + \left(\frac{\alpha}{\alpha+\beta(1-r)}\right)y \tag{6}$$

Proof. See Appendix A.1. □

Replicating the Dispersed Information Structure. Proposition 1 is an aggregate-equivalence result in that it matches the entire hierarchy of average expectations and therefore the aggregate action, but it does not match the full cross-sectional distribution of individual actions. The difference in the distribution of information across the two aggregate-equivalent economies is stark. In the dispersed-information economy, everyone is symmetrically imperfectly informed. In the hierarchical economy, by contrast, informed agents see the true fundamental while uninformed observe only a public signal. One can nevertheless refine Proposition 1 so that aggregate equivalence with MS remains exact *and* the type-specific individual action can be made arbitrarily close to the original MS individual action as shown by the following theorem.

Theorem 1 (Russian Doll Representation)**.** *Fix any integer $J \geq 1$ and partition the population into types $j = 0, 1, \dots, J$ with shares $p_j > 0$ and $\sum_{j=0}^{J} p_j = 1$. Type j observes the public signal y and the private signals*

$$s_{i,\ell}(\delta) = \theta + \nu_{i,\ell}(\delta), \qquad \nu_{i,\ell}(\delta) \sim N\big(0, \Delta_\ell(\delta)^{-1}\big), \qquad \ell = 0, \dots, j,$$

where the shocks are independent across agents, information layers, and (θ, y). The information sets, defined by $\Omega_{i,j}(\delta) = \sigma\big(y, s_{i,0}(\delta), \dots, s_{i,j}(\delta)\big)$, are thus strictly nested $\Omega_{i,0}(\delta) \subsetneq \Omega_{i,1}(\delta) \subsetneq \cdots \subsetneq \Omega_{i,J}(\delta)$, and there exist $\bar{\delta}_J > 0$ and information structures indexed by $\delta \in (0, \bar{\delta}_J)$ such that

(i) Exact aggregate equivalence. *For every $\delta \in (0, \bar{\delta}_J)$ the aggregate equilibrium action coincides exactly with the MS aggregate*

$$\bar{a} = y + \frac{\mu(1-r)}{1-r\mu}(\theta - y)$$

(ii) Higher-order beliefs. *Let $\overline{\mathbb{E}}_\delta$ denote the population-average conditional-expectation operator generated by the nested information structure. The entire hierarchy of average expectations coincides with its MS counterpart*

$$\overline{\mathbb{E}}_\delta^k[\theta] = y + \mu^k(\theta - y), \qquad k \geq 1. \tag{7}$$

(ii) Individual approximation. *Let $a_{i,j}(\delta)$ be the equilibrium action of a type-j agent and a_i^{MS} the equilibrium action in the original MS economy. As $\delta \downarrow 0$,*

$$\max_{0 \leq j \leq J} \Big| \mathbb{E}\big[a_{i,j}(\delta) \mid \theta, y\big] - \mathbb{E}\big[a_i^{\text{MS}} \mid \theta, y\big] \Big| \to 0$$
$$\max_{0 \leq j \leq J} \Big| \text{Var}\big(a_{i,j}(\delta) \mid \theta, y\big) - \text{Var}\big(a_i^{\text{MS}} \mid \theta, y\big) \Big| \to 0$$

Because the equilibrium actions are conditionally Gaussian, their conditional distributions therefore converge to the MS conditional distribution uniformly across the finitely many types.

Proof. The first step in the proof maps the incremental signal precisions $\Delta_0(\delta), \dots, \Delta_J(\delta)$ into implied posterior gains. Toward that end, let $\bar{j}_p \equiv \sum_{j=0}^{J} p_j j \in (0, J)$ denote the population-weighted average type.

For $\delta > 0$, the gain schedule $g_j(\delta) = \mu + \delta\big(j - \bar{j}_p\big)$ is the simplest increasing schedule that preserves the average gain

$$g_{j+1}(\delta) - g_j(\delta) = \delta > 0 \qquad \sum_{j=0}^{J} p_j g_j(\delta) = \mu \tag{8}$$

where the parameter δ measures the informational separation between adjacent types. Because the gains are increasing, they all lie in $(0,1)$ if and only if the following two endpoint conditions hold $g_0(\delta) = \mu - \delta \bar{j}_p > 0$, and $g_J(\delta) = \mu + \delta(J - \bar{j}_p) < 1$. Thus define

$$\bar{\delta}_J \equiv \min\left\{ \frac{\mu}{\bar{j}_p}, \frac{1-\mu}{J - \bar{j}_p} \right\}$$

where $\bar{\delta}_J$ is simply the largest admissible upper bound on the gain spacing.

We now back out the corresponding signal precisions associated with this gain schedule. The cumulative private precision that generates gain $g_j(\delta)$ is given by

$$B_j(\delta) = \frac{\alpha g_j(\delta)}{1 - g_j(\delta)}$$

and define $\Delta_0(\delta) = B_0(\delta)$, $\Delta_j(\delta) = B_j(\delta) - B_{j-1}(\delta)$, for $j = 1, \ldots, J$. Because the mapping from g to $\alpha g/(1-g)$ is strictly increasing, $\Delta_j(\delta) > 0$ for every $\delta \in (0, \bar{\delta}_J)$. The resulting information sets are therefore strictly nested, and their implied posterior gains are given by

$$\frac{B_j(\delta)}{\alpha + B_j(\delta)} = g_j(\delta)$$

The precision-weighted private signal of type j is therefore $x_{i,j}(\delta) = \frac{1}{B_j(\delta)} \sum_{\ell=0}^{j} \Delta_\ell(\delta) s_{i,\ell}(\delta) \sim N\big(\theta, B_j(\delta)^{-1}\big)$, with posterior mean of θ given by $m_{i,j}(\delta) = y + g_j(\delta)\big(x_{i,j}(\delta) - y\big)$.

To establish aggregate equivalence, conjecture $\bar{a} = y + q(\theta - y)$. The type-$j$ best response is then

$$a_{i,j}(\delta) = y + \big[(1-r) + rq\big]\big(m_{i,j}(\delta) - y\big) \tag{9}$$

Conditional on (θ, y), idiosyncratic signal errors average to zero within each type. Using (8), aggregation and equilibrium consistency therefore imply

$$q = \big[(1-r) + rq\big] \sum_{j=0}^{J} p_j g_j(\delta) = \mu\big[(1-r) + rq\big]$$

and therefore $q = \frac{\mu(1-r)}{1-r\mu}$, which is exactly the MS aggregate coefficient and is independent of δ. The equilibrium is unique because $r\mu < 1$. This proves part (i).

The same average-gain restriction preserves the entire hierarchy of average expectations. Let $H_0 = \theta$,

$H_k = \overline{\mathbb{E}}_\delta[H_{k-1}]$ $k \geq 1$. Suppose that $H_{k-1} = y + \mu^{k-1}(\theta - y)$, then

$$H_k = y + \mu^{k-1}\overline{\mathbb{E}}_\delta[\theta - y] = y + \mu^{k-1}\left(\sum_{j=0}^{J} p_j g_j(\delta)\right)(\theta - y) = y + \mu^k(\theta - y)$$

Since the result holds for $H_0 = \theta$, induction gives $\overline{\mathbb{E}}_\delta^k[\theta] = y + \mu^k(\theta - y)$ $k \geq 1$, which is equivalent to the MS economy and proves part (ii).

Finally, define $\kappa \equiv (1-r)/(1-r\mu)$ and substituting into (9) gives $a_{i,j}(\delta) = y + \kappa g_j(\delta)\big(x_{i,j}(\delta) - y\big)$. Conditional on (θ, y),

$$\mathbb{E}\left[a_{i,j}(\delta) \mid \theta, y\right] = y + \kappa g_j(\delta)(\theta - y) \qquad \operatorname{Var}\big(a_{i,j}(\delta) \mid \theta, y\big) = \frac{\kappa^2 g_j(\delta)^2}{B_j(\delta)} = \frac{\kappa^2}{\alpha} g_j(\delta)\left[1 - g_j(\delta)\right]$$

The corresponding MS moments are obtained by replacing $g_j(\delta)$ with μ. Since

$$\max_{0 \leq j \leq J} \left| g_j(\delta) - \mu \right| \leq J\delta \longrightarrow 0,$$

the conditional means and variances converge uniformly to their MS counterparts as $\delta \downarrow 0$. Given Gaussianity of the shocks, this proves part (iii).

Moreover, $\Delta_0(\delta) \longrightarrow \beta$, $\Delta_j(\delta) \longrightarrow 0$, for $j = 1, \ldots, J$. Thus, every hierarchy is strictly nested when $\delta > 0$, while in the limit its common first layer becomes the original MS private signal and its additional information layers vanish. □

The empirical role of hierarchy depth. Theorem 1 requires only $J = 1$. Thus, two strictly nested information types are sufficient to reproduce the MS aggregate exactly while approximating MS individual behavior arbitrarily closely. However, the parameter J provides needed flexibility to match the cross-sectional distribution of various statistics.

To see this, fix the MS parameters (α, β, r) and let $\mu \equiv \beta/(\alpha + \beta)$. For a hierarchy with types $j = 0, \ldots, J$, consider population shares and Kalman gains satisfying $p_j > 0$, $\sum_{j=0}^{J} p_j = 1$, $0 < g_0 < g_1 < \cdots < g_J < 1$ and $\sum_{j=0}^{J} p_j g_j = \mu$. Denote the set of all such partitions by $\Theta_J(\mu)$. Every partition in $\Theta_J(\mu)$ is consistent with Theorem 1, which identifies aggregate behavior but not the population shares, the type-specific gains, or the number of information types. These objects can be distinguished only with additional information about the cross-sectional distribution of behavior. This observation explains the empirical role of J; increasing J creates additional degrees of freedom needed to match cross-sectional features of actions, beliefs, or forecast errors. The following corollary states this formally.

Corollary 1 (Cross-Sectional Moment Matching)**.** *The set of aggregate-equivalent partitions $\Theta_J(\mu)$ has $2J$ degrees of freedom. Let*

$$\mathcal{M}_J : \Theta_J(\mu) \longrightarrow \mathbb{R}^q$$

collect q moments of the cross-sectional distribution of individual actions, beliefs, or forecast errors. Suppose that these moments respond independently to changes in the partition at some $\bar{\vartheta} \in \Theta_J(\mu)$, in the sense *that*

$$\operatorname{rank} D\mathscr{M}_J(\bar{\vartheta}) = q. \tag{10}$$

Then every vector of moments sufficiently close to $\mathscr{M}_J(\bar{\vartheta})$ can be matched exactly by a nearby partition in $\Theta_J(\mu)$. A necessary condition for matching q moments is

$$J \geq \left\lceil \frac{q}{2} \right\rceil \tag{11}$$

or, equivalently, there exist at least $\lceil q/2 \rceil + 1$ distinct information types.

Proof. See Appendix A.2 □

A simple numerical illustration. Let $\alpha = \beta = 1$ and $r = 1/2$, then $\mu = \beta/(\alpha+\beta) = 1/2$ and $\kappa = (1-r)/(1-r\mu) = 2/3$. The Morris–Shin aggregate is $\bar{a}^{\mathrm{MS}} = y + \frac{1}{3}(\theta - y)$. To illustrate Theorem 1, consider two equally sized types ($p_0 = p_1 = 1/2$) with gains $g_0 = \frac{1}{2} - \frac{\delta}{2}$, $g_1 = \frac{1}{2} + \frac{\delta}{2}$ and set $\delta = 1/5$. This yields $g_0 = 2/5$ and $g_1 = 3/5$, whose weighted average is exactly $\mu = 1/2$. The associated cumulative precisions are $B_j = \alpha g_j/(1-g_j)$, so $B_0 = 2/3$ and $B_1 = 3/2$. Type 0 therefore observes a single private signal of precision $2/3$; type 1 observes an additional refinement of precision $\Delta_1 = 5/6$. Equilibrium actions are $a_{i,j} = y + \kappa g_j(x_{i,j} - y)$. The conditional mean coefficients are $\kappa g_0 = 4/15$ and $\kappa g_1 = 2/5$; their population average is exactly $1/3$. Individual behavior differs in that each type's conditional variance equals $\frac{4}{9} g_j(1-g_j) = \frac{1}{9}(1-\delta^2) = 8/75$, whereas the original MS variance is $1/9$. Both the mean coefficients and the variances converge to their MS counterparts as $\delta \downarrow 0$, while the aggregate coefficient remains exactly $1/3$ for every admissible δ.

The same example illustrates Corollary 1. Parameterize the two-type hierarchy by

$$p_0 = p, \quad p_1 = 1-p, \qquad g_0 = \tfrac{1}{2} - (1-p)d, \qquad g_1 = \tfrac{1}{2} + pd \tag{12}$$

with $d > 0$ small. The average-gain restriction continues to hold for any (p, d), so the aggregate is invariant. Normalizing $\theta - y = 1$ and writing $\tilde{a}_j = \mathbb{E}[a_{i,j} \mid \theta, y]$, the cross-sectional variance and third central moment become

$$\mathscr{V} = \tfrac{4}{9} p(1-p)d^2 \qquad \mathscr{S} = \tfrac{8}{27} p(1-p)(2p-1)d^3$$

At the point $(p, d) = (1/2, 1/5)$ one obtains $\mathscr{V} = 1/225$ and $\mathscr{S} = 0$. The Jacobian matrix of $(\mathscr{V}, \mathscr{S})$ with respect to (p, d) evaluated at this point is

$$\begin{pmatrix} 0 & 2/45 \\ -4/3375 & 0 \end{pmatrix}$$

which has nonzero determinant. Consequently two free parameters can match any nearby targets for cross-sectional variance and skewness while the aggregate action remains exactly $y + \frac{1}{3}(\theta - y)$.

Implications. The broader implication is that our "Russian doll" information structure can reproduce the MS aggregate, its higher-order beliefs, and individual actions. Our results clarify what aggregate data can and cannot reveal in these canonical models. Aggregate behavior identifies μ, but cannot distinguish symmetric dispersed information from a hierarchy of information types. Cross-sectional evidence is therefore essential for learning about the *distribution* of information. Hierarchical depth supplies the additional degrees of freedom needed to match such evidence, while the modeler can choose the relevant moments and estimate the resulting partition. These insights extend to dynamic environments as we now demonstrate.

2.2 DYNAMIC MODEL

We now examine the dynamic Beauty Contest model as presented by Angeletos and Huo (2021) (AH hereafter). We first briefly review their results and note that we employ the original notation of their paper, so β is no longer the Morris and Shin (2002) precision.

Time is discrete with a continuum of agents indexed by $i \in [0,1]$ choosing action $a_{i,t} \in \mathbb{R}$. Let $a_t = \int_0^1 a_{i,t}\, di$ denote the average action. The best-response function takes the form $a_{i,t} = \mathbb{E}_{i,t}\left[\varphi \xi_t + \beta a_{i,t+1} + \gamma a_{t+1}\right]$ where ξ_t is an underlying fundamental, $\mathbb{E}_{i,t}[\cdot]$ is agent i's conditional expectation at date t, $\varphi > 0$ scales exposure to the fundamental, $\delta \in (0,1)$ is the discount factor, and where $\beta = \delta - \gamma$, $\gamma \in [0,\delta)$ measures the strength of strategic interactions. The fundamental is assumed to follow an AR(1) process

$$\xi_t = \rho \xi_{t-1} + \eta_t = \left(\frac{1}{1-\rho L}\right)\eta_t \qquad \eta_t \sim \mathcal{N}(0,1)$$

with persistence $\rho \in (0,1)$ and where L is the lag operator, $Lx_t = x_{t-1}$. Each agent i receives a private signal in every period $x_{i,t} = \xi_t + u_{i,t}$ where $u_{i,t} \sim \mathcal{N}(0,\sigma_u^2)$. Agent i's information set at t is the history of their private signals up to date t, $\Omega_t^i = \{x_{i,t}, x_{i,t-1}, \ldots\}$.

AH show that the first-order average belief follows an AR(2) process in the η_t space and AR(1) in ξ_t

$$\overline{\mathbb{E}}_t[\xi_t] = \left(1 - \frac{\lambda}{\rho}\right)\left(\frac{1}{1-\lambda L}\right)\xi_t \tag{13}$$

where $\lambda = \rho(1-g)$ and $g \in (0,1)$ is the steady-state Kalman gain (decreasing in σ_u^2). Higher-order beliefs $\overline{\mathbb{E}}_t[\overline{\mathbb{E}}_{t+1}[\cdots[\overline{\mathbb{E}}_{t+h-1}[\xi_{t+h}]]]$ follow ARMA processes of ever-increasing order, generating a "curse of dimensionality" in the state space (Angeletos and Huo, 2021, p. 11). Despite this complexity, Proposition 2 (Angeletos and Huo, 2021, eqs. (16)–(17)) demonstrates that the *aggregate* rational-expectations equilibrium is an AR(2)

$$a_t = \left(1 - \frac{\vartheta}{\rho}\right)\left(\frac{1}{1-\vartheta L}\right)a_t^* \qquad\qquad a_t^* = \left(\frac{\varphi}{1-\rho\delta}\right)\xi_t \tag{14}$$

where a_t^* is the full-information rational expectations solution and $\vartheta \in (0,\rho)$ is the reciprocal of the largest root of the cubic

$$C(z) \equiv -z^3 + \left(\rho + \tfrac{1}{\rho} + \tfrac{1}{\rho\sigma_u^2} + (\delta-\gamma)\right)z^2 - \left(1 + (\delta-\gamma)\left(\rho + \tfrac{1}{\rho}\right) + \tfrac{\delta}{\rho\sigma_u^2}\right)z + (\delta-\gamma) \tag{15}$$

When $\gamma = 0$, $\vartheta = \lambda$ and the equilibrium is determined by first-order beliefs. When $\gamma > 0$, $\lambda < \vartheta < \rho$, strategic interactions make the equilibrium less responsive on impact and more persistent than the first-order average belief.

Dispersed–Hierarchical Equivalence. The static Russian-Doll construction of Theorem 1 exploits the fact that, conditional on the public signal, common uncertainty is one-dimensional and can be characterized by a scalar. In this model, the restriction $\sum_j p_j g_j = \mu$ is therefore sufficient to match population-average expectations. In the dynamic beauty contest model, the equivalent sufficient statistic is a causal forecasting filter. Simply adding shocks across types (as in the static case) would produce type-specific filtering roots and, in turn, would alter the persistence of the aggregate equilibrium. The discussion below establishes, via reverse-engineering, an information structure that reproduces the AH aggregate AR(2) exactly while approximating the corresponding individual actions arbitrarily closely and produces a dynamic equivalent of Theorem 1.

Toward that end, define the average expectations in the dispersed-information AH economy for any covariance stationary variable Z by $\overline{\mathbb{E}}_t^{\text{AH}}[Z] \equiv \int_0^1 \mathbb{E}[Z \mid x_i^t]\, di$, with $x_i^t \equiv \{x_{i,t}, x_{i,t-1}, \ldots\}$ such that $\mathscr{C} \equiv \overline{\text{span}}\{\eta_s : s \in \mathbb{Z}\}$ in the standard L^2 norm with finite second moments. Iterating the individual best response forward and aggregating gives

$$a_t = \overline{\mathbb{E}}_t^{\text{AH}}\left[\left(\frac{\varphi}{1-\beta L^{-1}}\right)\xi_t + \left(\frac{\gamma L^{-1}}{1-\beta L^{-1}}\right)a_t\right]$$

Conjecture that the aggregate action has moving-average representation $a_t = A(L)\eta_t$, and define

$$G_A(L) \equiv \frac{\varphi}{(1-\beta L^{-1})(1-\rho L)} + \frac{\gamma L^{-1} A(L)}{1-\beta L^{-1}}$$

Thus, the equilibrium restriction is given by $A(L)\eta_t = \overline{\mathbb{E}}_t^{\text{AH}}\left[G_A(L)\eta_t\right]$. When $\gamma = 0$, equality of first-order average beliefs is sufficient for aggregate equivalence but when $\gamma > 0$, first-order matching is no longer sufficient. Thus, we are looking for a fixed-point problem in both functional form and expectation.

Our hierarchical economy seeks a partition into types $j = 0, \ldots, J$, with type sets $\mathscr{I}_j$ and shares p_j. For a candidate type-j signal history $x_{i,j}^t$, define $\overline{\mathbb{E}}_{j,t}[Z] \equiv \frac{1}{p_j}\int_{\mathscr{I}_j} \mathbb{E}[Z \mid x_{i,j}^t]\, di$. Now suppose type-specific signals satisfied the proportional restriction

$$\overline{\mathbb{E}}_{j,t}[Z] = c_j \overline{\mathbb{E}}_t^{\text{AH}}[Z] \tag{16}$$

with $Z \in \mathscr{C}$. If $\sum_{j=0}^J p_j c_j = 1$, then population-average expectations in the hierarchical economy would satisfy

$$\overline{\mathbb{E}}_t[Z] \equiv \sum_{j=0}^J p_j \overline{\mathbb{E}}_{j,t}[Z] = \sum_{j=0}^J p_j c_j \overline{\mathbb{E}}_t^{\text{AH}}[Z] = \overline{\mathbb{E}}_t^{\text{AH}}[Z]$$

Moreover, one could match individual actions. Let $A^\star(L)$ denote the AH solution to $\overline{\mathbb{E}}_t^{\text{AH}}[G_A(L)\eta_t]$, and define the type-specific average action by $a_{j,t} \equiv p_j^{-1}\int_{\mathscr{I}_j} a_{i,t}\, di$. Enforcing $\overline{\mathbb{E}}_{j,t}[Z] = c_j \overline{\mathbb{E}}_t^{\text{AH}}[Z]$ yields

$$a_{j,t} = c_j \overline{\mathbb{E}}_t^{\text{AH}}\left[G_{A^\star}(L)\eta_t\right] = c_j A^\star(L)\eta_t \tag{17}$$

Every type-average action therefore has the same moving-average shape as the AH aggregate, scaled by c_j, and these scales average to one. In particular, if the AH solution has reduced denominator $(1-\rho L)(1-\vartheta L)$, the hierarchical economy inherits the same aggregate AR(2) and introduces no additional persistence roots. The solution algorithms, equilibrium restrictions and proofs follow immediately.

What information set is consistent with our proportionality constraint (16)? Let the effective signal of type j be $x_{i,j,t} = \xi_t + u_{i,j,t}$, where $u_{i,j,t}$ is independent of the fundamental and has spectral density $S_{u,j}(z)$. Let $S_{x,j}(z) = B_j(z)B_j(z^{-1})$ denote the unique canonical spectral factorization of the type-j signal spectrum. Because $S_{Zx_j}(z) = S_{Z\xi}(z)$ for every $Z \in \mathscr{C}$, the Wiener–Kolmogorov formula and the cross-sectional law of large numbers give

$$\overline{\mathbb{E}}_{j,t}[Z] = \frac{1}{B_j(L)}\left[\frac{S_{Z\xi}(L)}{B_j(L^{-1})}\right]_+ \xi_t$$

Choosing $B_j(z) = \frac{B(z)}{\sqrt{c_j}}$ therefore yields $\overline{\mathbb{E}}_{j,t}[Z] = c_j\overline{\mathbb{E}}_t^{\text{AH}}[Z]$, with $S_{x,j}(z) = \frac{S_x(z)}{c_j}$ Since $x_{i,j,t} = \xi_t + u_{i,j,t}$, the required measurement-noise spectrum is therefore

$$S_{u,j}(z) = \frac{S_x(z)}{c_j} - S_\xi(z) = \frac{\sigma_u^2}{c_j} + \left(\frac{1}{c_j} - 1\right) S_\xi(z) \tag{18}$$

The measurement noise is colored whenever $c_j \neq 1$. Because $\sum_j p_j c_j = 1$, any nondegenerate hierarchy contains types with $c_j < 1$ (a less-informed type who reacts more slowly to signals) and types with $c_j > 1$ (a more informed type who reacts more aggressively to signals), although no type need have $c_j = 1$. For every nonzero lag h,

$$\operatorname{Cov}\big(u_{i,j,t}, u_{i,j,t-h}\big) = \left(\frac{1}{c_j} - 1\right)\operatorname{Cov}\big(\xi_t, \xi_{t-h}\big). \tag{19}$$

The role of this covariance restriction is most easily understood by comparison with the static model. Because the static model contains no signal history, its filtering problem is algebraically analogous to setting $\lambda = 0$ in the dynamic kernel $(1-\lambda L)^{-1}$. Information types can then differ through their contemporaneous signal variances alone. When $\lambda \neq 0$, however, agents use the entire signal history. Changing only the contemporaneous noise variance alters the signal covariance at lag zero while leaving its nonzero-lag covariances unchanged, thereby changing the relative covariance structure of the signal and producing type-specific filtering roots λ_j.

Equation (19) supplies exactly the missing nonzero-lag covariances. Together with the corresponding adjustment at lag zero, it ensures $\operatorname{Cov}\big(x_{i,j,t}, x_{i,j,t-h}\big) = \frac{1}{c_j}\operatorname{Cov}\big(x_{i,t}, x_{i,t-h}\big)$ for every h. Thus, the entire signal covariance function—not merely its variance—is rescaled by $1/c_j$, leaving the filtering root λ unchanged. When $c_j < 1$, positive serial correlation enlarges the signal autocovariances proportionally. When $c_j > 1$, negative serial correlation offsets part of the persistence inherited from the fundamental. Equivalently, although the effective measurement noises $u_{i,j,t}$ are colored, the incremental layers $\zeta_{i,\ell,t}$ become white after applying the common prewhitening filter $B(L)^{-1}$. The dynamic construction is therefore the static Gaussian refinement applied innovation by innovation, with precisely the temporal correlation required to preserve the AH causal filter. We have established the following result.

Theorem 2 (Dynamic Russian-Doll Representation)**.** *Consider the dispersed-information economy of An-*

geletos and Huo (2021) and let $S_\xi(z)$ and $S_x(z) = S_\xi(z) + \sigma_u^2$ denote the corresponding spectra, and define $\bar{c} \equiv 1 + \sigma_u^2(1-\rho)^2 > 1$. Fix any integer $J \ge 1$ and define the set of admissible hierarchical partitions

$$\Theta_J^{\mathrm{AH}} \equiv \{(p,c) : p_j > 0, \quad \sum_{j=0}^{J} p_j = 1, \quad 0 < c_0 < \cdots < c_J < \bar{c}, \quad \sum_{j=0}^{J} p_j c_j = 1\} \tag{20}$$

For any $(p,c) \in \Theta_J^{\mathrm{AH}}$, partition the population into sets $\mathscr{I}_j$, $j = 0,\ldots,J$, with shares p_j. Let $u_{i,J}$ and $\zeta_{i,1},\ldots,\zeta_{i,J}$ be stationary Gaussian processes that are mutually independent across layers, independent across agents, and independent of the fundamental, with spectra

$$S_{u,J}(z) = \frac{S_x(z)}{c_J} - S_\xi(z), \qquad S_{\zeta,\ell}(z) = S_x(z)\left(\frac{1}{c_{\ell-1}} - \frac{1}{c_\ell}\right), \qquad \ell = 1,\ldots,J.$$

Construct the effective signals recursively by $x_{i,J,t} = \xi_t + u_{i,J,t}$, $x_{i,j-1,t} = x_{i,j,t} + \zeta_{i,j,t}$, $j = 1,\ldots,J$ and let type j observe

$$\Omega_{i,j,t} \equiv \sigma\big(x_{i,0}^t,\ldots,x_{i,j}^t\big), \qquad x_{i,j}^t \equiv \{x_{i,j,t}, x_{i,j,t-1},\ldots\}. \tag{21}$$

These information sets are strictly nested, $\Omega_{i,0,t} \subsetneq \Omega_{i,1,t} \subsetneq \cdots \subsetneq \Omega_{i,J,t}$. Define the within-type and population-average expectations by

$$\overline{\mathbb{E}}_{j,t}[Z] \equiv \frac{1}{p_j}\int_{\mathscr{I}_j} \mathbb{E}\big[Z \mid \Omega_{i,j,t}\big]\, di \qquad \overline{\mathbb{E}}_t[Z] \equiv \sum_{j=0}^{J} p_j \overline{\mathbb{E}}_{j,t}[Z]$$

Then the following results hold.

(i) Exact average-expectation and higher-order-belief equivalence. *For every $Z \in \mathscr{C}$, every date t, and every type j,*

$$\overline{\mathbb{E}}_{j,t}[Z] = c_j \overline{\mathbb{E}}_t^{\mathrm{AH}}[Z] \qquad \overline{\mathbb{E}}_t[Z] = \overline{\mathbb{E}}_t^{\mathrm{AH}}[Z] \tag{22}$$

Consequently, the entire hierarchy of population-average fixed-date, time-shifted, and mixed-date higher-order beliefs coincides with its dispersed-information AH counterpart.

(ii) Exact aggregate equivalence. *Suppose $0 \le \beta < 1$, $\gamma \ge 0$, and $\beta + \gamma < 1$. For every $(p,c) \in \Theta_J^{\mathrm{AH}}$, the unique bounded square-integrable aggregate equilibrium satisfies $a_t = a_t^{\mathrm{AH}}$, where a_t^{AH} is the Angeletos–Huo AR(2) in (14).*

(iii) Individual approximation. *Fix the shares p, let $\bar{j}_p \equiv \sum_{j=0}^{J} p_j j$, and define*

$$c_j(\varepsilon) \equiv 1 + \varepsilon(j - \bar{j}_p) \qquad \bar{\varepsilon}_J \equiv \min\left\{\frac{1}{\bar{j}_p}, \frac{\bar{c}-1}{J - \bar{j}_p}\right\} \tag{23}$$

For every $\varepsilon \in (0,\bar{\varepsilon}_J)$, $(p, c(\varepsilon)) \in \Theta_J^{\mathrm{AH}}$. Let $a_{i,j,t}(\varepsilon)$ denote the resulting equilibrium action of a type-j agent, let $a_{i,t}^{\mathrm{AH}}$ denote an individual action in the original dispersed-information economy, and *define* $\mathscr{F}_t^{\xi} \equiv \sigma(\xi_s : s \leq t)$. *As* $\varepsilon \downarrow 0$,

$$\max_{0\leq j\leq J}\left|\mathbb{E}\left[a_{i,j,t}(\varepsilon) \mid \mathscr{F}_t^{\xi}\right] - \mathbb{E}\left[a_{i,t}^{\mathrm{AH}} \mid \mathscr{F}_t^{\xi}\right]\right| \longrightarrow 0 \tag{24}$$

$$\max_{0\leq j\leq J}\left|\mathrm{Var}\left(a_{i,j,t}(\varepsilon) \mid \mathscr{F}_t^{\xi}\right) - \mathrm{Var}\left(a_{i,t}^{\mathrm{AH}} \mid \mathscr{F}_t^{\xi}\right)\right| \longrightarrow 0 \tag{25}$$

The conditional-mean convergence holds almost surely and in L^2. Because the actions are conditionally Gaussian, their date-t conditional distributions converge weakly to the AH conditional distribution, uniformly across the finitely many types.

Proof. The grid construction (23) and proof of convergence in individual actions follows the intuition of the static case and so we relegate a complete proof to Appendix A.3. □

2.3 IMPLICATIONS. Theorem 2 is more than just an observational equivalence as there are two important implications that address current challenges in the literature. First, higher-order beliefs are the central computational obstacle in dynamic incomplete-information models. Angeletos and Huo (2021) characterize these beliefs explicitly when signals are exogenous, and Huo and Takayama (2025) show more generally that the equilibrium admits a finite-state representation whenever signals follow finite ARMA processes. However, Huo and Takayama (2025) note that the finite-state result may fail once information is endogenous, and Chahrour and Jurado (2025) prove that in the original model of Townsend (1983) (with a noisy price index) is infinite-dimensional, so the endogenous processes admit no finite ARMA representation. Our Russian-Doll economies make higher-order belief calculations straightforward, even when information is endogenous as we show in Section 3. Second, it is well known that identifying the underlying information structure from aggregate outcomes is challenging (Bergemann and Morris, 2013). Our equivalence result demonstrates that crucial information frictions, e.g., Coibion and Gorodnichenko (2015), live in an equivalence class with a resolution that comes from cross sectional information only. Such informational variation does not exist when signals are symmetrically distributed.

Higher-order beliefs. Higher-order beliefs are a central computational obstacle in dynamic incomplete-information models, especially when information is endogenous. While our results here mirror AH, the true payoff comes in the next section (Section 3) where information is endogenous.

The static Russian-Doll model provides the intuition. The payoff-relevant uncertainty is one-dimensional in that one average expectation multiplies the fundamental by the scalar gain μ, and an order-m belief therefore multiplies it by μ^m. Dynamics preserve this logic but make the payoff-relevant uncertainty multidimensional. The object being forecast is now a vector containing the effects of innovations of different ages, and the scalar gain μ is replaced by a single matrix $\boldsymbol{K}$. One average expectation multiplies the vector by $\boldsymbol{K}$; an order-m fixed-date belief multiplies it by $\boldsymbol{K}^m$.

To see this cleanly, define $\boldsymbol{\eta}_t \equiv \left(\eta_t, \eta_{t-1}, \eta_{t-2}, \ldots\right)^{\top}$ and write any admissable linear combination as $Z_t = \boldsymbol{a}'\boldsymbol{\eta}_t = \sum_{n=0}^{\infty} a_n \eta_{t-n}$, where $\boldsymbol{a} \equiv (a_0, a_1, a_2, \ldots)^{\top}$. A direct projection onto the orthogonal signal innovations constructed as above gives $\overline{\mathbb{E}}_{j,t}[Z_t] = c_j\big(\boldsymbol{K}\boldsymbol{a}\big)'\boldsymbol{\eta}_t$, where $\boldsymbol{K}$ has entries

$$K_{kn} = \chi \sum_{\ell=0}^{\min\{k,n\}} \lambda^{k+n-2\ell} = \frac{\chi}{1-\lambda^2}\left(\lambda^{|k-n|} - \lambda^{k+n+2}\right) \qquad \chi \equiv \mu(1-\rho\lambda) \tag{26}$$

The row index k records the age of an innovation in the resulting belief, while the column index n records its age in the object being forecast. The first few entries are given by

$$\boldsymbol{K} = \chi \begin{pmatrix} 1 & \lambda & \lambda^2 & \lambda^3 & \cdots \\ \lambda & 1+\lambda^2 & \lambda+\lambda^3 & \lambda^2+\lambda^4 & \cdots \\ \lambda^2 & \lambda+\lambda^3 & 1+\lambda^2+\lambda^4 & \lambda+\lambda^3+\lambda^5 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \end{pmatrix} \tag{27}$$

This matrix is the dynamic analogue of the static posterior gain. When $\lambda = 0$, $\boldsymbol{K} = \mu\boldsymbol{I}$, and innovations of different ages are filtered independently, revealing the static scalar-gain formula for higher-order beliefs. When $\lambda > 0$, the off-diagonal entries capture the fact that a signal about the current fundamental confounds innovations of different ages. The matrix is nevertheless known explicitly, so this temporal confounding creates no new filtering problem at a higher belief order.

When expectations are formed at different dates, we can define $\boldsymbol{S}_d$ to denote the shift that deletes innovations that have not yet occurred and relabels the remaining coefficients relative to the current date: $\big(\boldsymbol{S}_d\boldsymbol{a}\big)_k \equiv a_{k+d}$. Thus, forming an expectation d periods before the date of a target requires two steps: first shift its coefficients by $\boldsymbol{S}_d$, and then apply $\boldsymbol{K}$. For example, consider a target dated T and expectation dates $\tau_1 \leq \cdots \leq \tau_m \leq T$. Set $\tau_{m+1} \equiv T$ and $d_r \equiv \tau_{r+1} - \tau_r$. For a sequence of information types $\boldsymbol{j} = (j_1, \ldots, j_m)$, the resulting belief is

$$\mathcal{H}_{\boldsymbol{j},\boldsymbol{\tau}}[Z_T] = \left(\prod_{r=1}^{m} c_{j_r}\right)\Big[\boldsymbol{K}\boldsymbol{S}_{d_1}\boldsymbol{K}\boldsymbol{S}_{d_2}\cdots\boldsymbol{K}\boldsymbol{S}_{d_m}\boldsymbol{a}\Big]'\boldsymbol{\eta}_{\tau_1}. \tag{28}$$

This formula contains the entire taxonomy summarized in Table 1. Fixed-date beliefs take powers of $\boldsymbol{K}$; consecutive-date beliefs alternate $\boldsymbol{K}$ with one-period shifts; mixed-date beliefs use the corresponding sequence of shifts; and information types enter only through the product of their scalar intensities. No new projection is solved at any stage.

The hierarchical information structure with proportional spectra implies that every agent type has the same matrix $\boldsymbol{K}$ and differs only through the scalar c_j. Without this restriction, type j would have its own matrix $\boldsymbol{K}_j$, and an order-m belief would involve products of different filtering matrices. The Dynamic Russian-Doll Theorem implies that the resulting population-average beliefs coincide exactly with those in the baseline AH economy. For example, two applications of the belief matrix give

$$\big(\boldsymbol{K}^2\boldsymbol{\rho}\big)'\boldsymbol{\eta}_t = \frac{\mu^2(1-\rho\lambda^2 L)}{(1-\lambda^2)(1-\rho L)(1-\lambda L)^2}\eta_t, \tag{29}$$

which is precisely the second-order AH belief. More generally, $\boldsymbol{K}^m\boldsymbol{\rho}$ generates the ARMA$(m+1, m-1)$ representation without requiring a new Wiener–Kolmogorov calculation at order m. AH obtain similarly

Table 1: A Taxonomy of Higher-Order Beliefs

Belief	Expectation dates and target	Population coefficient vector
First-order, horizon h	$\overline{\mathbb{E}}_t[\xi_{t+h}]$	$\boldsymbol{K}\boldsymbol{S}_h\boldsymbol{\rho}$
Fixed-date, order m	$\overline{\mathbb{E}}_t[\cdots\overline{\mathbb{E}}_t[\xi_{t+h}]\cdots]$	$\boldsymbol{K}^m\boldsymbol{S}_h\boldsymbol{\rho}$
Consecutive-date, order m	$\overline{\mathbb{E}}_t[\cdots\overline{\mathbb{E}}_{t+m-1}[\xi_{t+m-1}]\cdots]$	$(\boldsymbol{K}\boldsymbol{S}_1)^{m-1}\boldsymbol{K}\boldsymbol{\rho}$
Mixed-date, order m	$\tau_1 \leq \cdots \leq \tau_m \leq T$	$\boldsymbol{K}\boldsymbol{S}_{d_1}\cdots\boldsymbol{K}\boldsymbol{S}_{d_m}\boldsymbol{\rho}$

Notes: $\boldsymbol{K}$ is the common Wiener–Kolmogorov matrix in the innovation basis; $\boldsymbol{S}_d$ is the d-period shift that deletes unrealized innovations and reindexes the remainder relative to the current date; $\boldsymbol{\rho} = (1, \rho, \rho^2, \ldots)^\top$ is the coefficient vector of the fundamental. Population-average beliefs are obtained by setting every intensity equal to one (or, equivalently, by averaging the type-labelled products $\prod_r c_{j_r}$). Type-labelled beliefs differ from the population objects only by the scalar product of the relevant intensities.

transparent innovation-age products in their Online Appendix H by replacing their baseline information structure with orthogonal signals about individual innovations; they note that exact observational equivalence is then lost. The Russian-Doll construction retains the baseline signal-extraction content in the off-diagonal entries of $\boldsymbol{K}$, while making the entire hierarchy a sequence of ordinary matrix multiplications. It therefore delivers both a tractable taxonomy of higher-order beliefs and exact equivalence with the original dispersed-information economy.

Identification. Theorem 2 shows that *every* partition in Θ_J^{AH} generates the same aggregate action, the same consensus forecasts, and the same aggregate dynamics. We now show that this aggregate equivalence does not imply that the information structure is unidentifiable. As in Corollary 1 identification comes from the cross-sectional distribution of information. However, it is a challenge for interpreting aggregate representations of the data.

Angeletos and Huo (2021, Proposition 3) interpret the equilibrium as a representative-agent economy with two behavioral distortions,

$$a_t = \varphi\xi_t + \delta\omega_f\,\mathbb{E}_t[a_{t+1}] + \omega_b\,a_{t-1} \tag{30}$$

$$\omega_f = \frac{\rho^2\delta - \vartheta}{\delta(\rho^2 - \vartheta^2)}, \qquad \omega_b = \frac{\vartheta\rho^2(1-\delta\vartheta)}{\rho^2 - \vartheta^2}, \tag{31}$$

where $\vartheta \in (0, \rho)$ solves (15), and $\omega_f < 1$ is a myopic distortion and $\omega_b > 0$ is an anchoring distortion. Follow up propositions show that both stronger when noise is larger or when general-equilibrium feedback γ is larger holding δ fixed. The coefficients depend on the information environment only through the aggregate root ϑ, which Theorem 2 holds fixed across the equivalence class. Thus, every partition delivers the same (ω_f, ω_b), the same consensus forecasts at every horizon, and the equilibrium. In other words, Theorem 2 offers an alternative reading of the same two coefficients with myopia and anchoring measuring the sluggishness of the resulting average response rather than a behavioral one.

The same logic applies to the statistic that has become the standard empirical measure of informational frictions. Coibion and Gorodnichenko (2015) regress the average forecast error on the average forecast revision and interpret a positive coefficient as evidence of sluggish information, since under full information the revision is known at the time the forecast is made and the coefficient is zero. The diagnostic is reduced form, but Angeletos and Huo (2021, Proposition 6) show that in this environment it has a closed-form structural counterpart,

$$K_{\text{CG}} = \lambda \left(\frac{\vartheta + \rho - \rho\vartheta(\lambda + \vartheta) - \rho\lambda\vartheta(1 - \lambda\vartheta)}{(\rho - \lambda)(1 - \lambda\vartheta)(\rho + \vartheta - \lambda\rho\vartheta)} \right) \tag{32}$$

a function of the Kalman gain through λ and of strategic interaction through ϑ. Again, because both λ and ϑ are held fixed across Θ_J^{AH}, every partition in the equivalence class produces the same K_{CG} at every horizon. Two economies in which information is distributed very differently—one with a small, very well informed group and one with mild heterogeneity throughout the population—generate identical CG regressions.

Notice that Theorem 2 implies second moments cannot help with the observational equivalence. To see this, consider an agent whose information is summarized by the loading c_j. The agent reacts c_j times as strongly as the consensus, which scales the variance by c_j^2, but does so on the basis of a signal whose entire spectrum is scaled by $1/c_j$. The net effect scales type-j variance by c_j, and since $\sum_j p_j c_j = 1$ the population average is unchanged. Formally, if $Y_{i,j,t} = c_j H(L) x_{i,j,t}$ for any common filter H, then $\text{Var}(Y_{i,j,t}) = c_j V_{\text{AH}}$ and $\sum_{j=0}^{J} p_j \text{Var}(Y_{i,j,t}) = V_{\text{AH}}$, where V_{AH} is the corresponding AH variance. Thus, variation in higher-order moments must be used to achieve identification but these are the exact moments that are fixed by a symmetrically dispersed information economy.

As in Corollary 1, identification comes from the cross-sectional variance of the information loadings,

$$v_c \equiv \text{Var}_p(c_j) = \sum_{j=0}^{J} p_j (c_j - 1)^2, \tag{33}$$

which measures how unequally information is distributed. It is zero in AH, where all agents share one signal precision, and positive in any nondegenerate hierarchy.

To see this more clearly, define cross-sectional dispersion at date t as $D_t \equiv \int_0^1 \left(a_{i,t} - a_t\right)^2 di$, which is the average squared distance between an individual action and the aggregate. In the symmetric AH economy all agents share the same signal precision, so dispersion is driven entirely by idiosyncratic noise and is identical in every period, at $D^{\text{AH}} = V_{\text{AH}} - V_a$ with $V_a \equiv \text{Var}(a_t)$ and $V_{\text{AH}} \equiv \text{Var}(a_{i,t})$. A hierarchy changes this because type-average actions are $c_j a_t$, so types fan out when the aggregate action is far from its mean and converge when it is near it. Dispersion now has two parts. The between-type part is $\sum_j p_j (c_j a_t - a_t)^2 = v_c a_t^2$, which moves with the state. The within-type part is constant at $D^{\text{AH}} - v_c V_a$ because better-informed types have less idiosyncratic noise, and the reduction is exactly $v_c V_a$ in population. Adding them, with actions written as deviations from unconditional means, gives

$$D_t = D^{\text{AH}} + v_c \left[a_t^2 - \mathbb{E}\left[a_t^2\right] \right] \tag{34}$$

Equation (34) is an exact population regression of dispersion on the squared aggregate action, with intercept D^{AH} and slope v_c. Regressing, say survey expectation dispersion on the squared aggregate response therefore estimates information inequality directly, and the model imposes the one-sided restriction $v_c \geq 0$.

Moreover, within a type, behavior is normal with variance $c_j V_{\text{AH}}$ but across types variances differ, producing a mixture of normals with a common mean and unequal variances (same variance as AH but positive excess kurtosis). Normalizing each even moment by its Gaussian counterpart removes the Gaussian shape and leaves exactly a raw moment of the loading distribution. Setting $m_0 \equiv 1$, for $q \geq 1$,

$$m_q \equiv \frac{\mathbb{E}\left[Y_{i,t}^{2q}\right]}{(2q-1)!!\, V_{\text{AH}}^q} = \sum_{j=0}^{J} p_j c_j^q, \qquad (2q-1)!! \equiv 1 \cdot 3 \cdots (2q-1), \tag{35}$$

so observable moments of behavior are observable moments of the information distribution.[2] Here $m_1 = 1$ is aggregate equivalence and $m_2 = 1 + v_c$ is information inequality, while higher m_q trace out the remaining shape. A simple counting of unknowns shows how many restrictions are needed. A distribution on $J+1$ unspecified points carries $2(J+1)$ unknowns—the loadings and the shares—and each moment supplies one equation, so $m_0, \ldots, m_{2J+1}$ is exactly enough. Two are free, since $m_0 = 1$ by construction and $m_1 = 1$ by aggregate equivalence, leaving $2J$ informative moments and matching the $2J$ degrees of freedom of Θ_J^{AH}. Thus, we have motivated the following theorem.

Theorem 3 (Global Identification from the Individual Distribution)**.** *Fix $(p, c) \in \Theta_J^{\text{AH}}$ and let $Y_{i,j,t} = c_j H(L) x_{i,j,t}$ be any centered individual action, belief, or forecast generated by a nonzero common filter, with pooled variance $V_{\text{AH}} > 0$, where moments average over agents and over the stationary distribution of aggregate histories. Then:*

(i) If the hierarchy has $J+1$ types, the moments $m_2, \ldots, m_{2J+1}$ determine the shares and the ordered loadings uniquely, and globally within Θ_J^{AH}.

(ii) If the number of types is unknown but finite, the full sequence $\{m_q\}_{q \geq 0}$ identifies it, along with the shares and loadings. Hence the stationary distribution of any one nondegenerate individual object identifies (J, p, c).

Proof. See Appendix A.4. □

The logic is the same as in the static case of Corollary 1. There, a partition had to satisfy $\sum_j p_j g_j = \mu$ as the gains could be spread across types however one liked, provided their population average matched the single gain of the symmetric economy. Here the requirement is $\sum_j p_j c_j = 1$, which is the same restriction written in relative terms, since $g_j = \mu c_j$ recovers the static form exactly. In both cases the partition has $2J$ free parameters after imposing the two adding-up conditions, so at most $2J$ independent cross-sectional moments can be matched.

[2] Conditional on type j, $Y_{i,j,t}$ is a linear filter of a Gaussian process, hence normal with variance $c_j V_{\text{AH}}$, and a mean-zero normal with variance s^2 satisfies $\mathbb{E}[Z^{2q}] = (2q-1)!!\, s^{2q}$. Averaging over types and dividing by $(2q-1)!!\, V_{\text{AH}}^q$ gives (35). Odd moments vanish because each type is symmetric about zero, so all the information sits in the even moments.

A simple numerical illustration. Let a share p of agents be better informed and $1-p$ less informed, with loadings separated by a single information gap $d > 0$:

$$c_0 = 1 - pd, \qquad c_1 = 1 + (1-p)d, \qquad p_0 = 1 - p, \quad p_1 = p. \tag{36}$$

The informed exceed the consensus by $(1-p)d$ and the uninformed fall short by pd, so $p_0 c_0 + p_1 c_1 = 1$ holds for any (p, d) and the aggregate is invariant. Take $p = 1/4$ and $d = 2/5$, so that a quarter of agents are informed. Then $c_0 = 9/10$ and $c_1 = 13/10$ and the uninformed react ten percent less strongly than the consensus and the informed thirty percent more. The loading distribution has variance and third central moment

$$v_c = p(1-p)d^2 = \tfrac{3}{100}, \qquad \tau_c = p(1-p)(1-2p)d^3 = \tfrac{3}{500}, \tag{37}$$

and by (35) these appear in the normalized moments of individual behavior as $m_2 = 1 + v_c = 103/100$ and $m_3 = 1 + 3v_c + \tau_c = 137/125$. Equivalently, the pooled distribution of individual actions has kurtosis $3(1 + v_c) = 3.09$ against 3 for the symmetric AH economy.

Both moments are needed for identification. Because $v_c = p(1-p)d^2$ is unchanged when p is replaced by $1-p$, the fourth moment alone cannot say whether the informed are a quarter of the population or three quarters. The sixth moment resolves this, since $\tau_c > 0$ when $p < 1/2$, $\tau_c < 0$ when $p > 1/2$, and $\tau_c = 0$ only at $p = 1/2$. Dividing out the scale gives a statistic in p alone,

$$\frac{\tau_c}{v_c^{3/2}} = \frac{1-2p}{\sqrt{p(1-p)}} = \tfrac{2}{\sqrt{3}} \approx 1.155, \tag{38}$$

which is strictly decreasing on $(0,1)$ from $+\infty$ to $-\infty$ and therefore inverts to a unique p; here it returns $p = 1/4$, and the gap follows from $d = [v_c/\{p(1-p)\}]^{1/2} = 2/5$. The fourth and sixth moments of individual behavior thus recover the entire information structure, while every aggregate object—the impulse responses, the consensus forecast at each horizon, K_{CG}, and the pooled forecast-error variance—is identical to AH.

3 ENDOGENOUS SIGNAL EXTRACTION

To quote Angeletos and Huo (2021), results derived thus far come “at the cost of some generality, in particular we abstract from the possible endogeneity of information.” We now extend the beauty contest model to allow for endogenous signal extraction and derive our equivalence result in this novel setting. Proposition 2 derives restrictions that yield an aggregate equivalence between the dispersed- and hierarchical-information economies. We then refine the two-type hierarchy to arbitrarily approximate a symmetric dispersed information economy in Theorem 4. We then study information transmission, a question that cannot be posed when information is exogenous.

3.1 DISPERSED-HIERARCHICAL EQUIVALENCE

Dispersed Information. We assume agents (indexed by i) observe the sequence of current and past aggregate actions $\{a_{t-j}\}_{j=0}^{\infty}$ in addition to a sequence of noisy signals about the fundamental innovation, $\eta_{it} = \eta_t + v_{it}$ with $v_{it} \sim \mathcal{N}(0, \sigma_v^2)$ for $i \in [0,1]$ and $\Omega_t^i = \{a_{t-j}, \eta_{i,t-j}\}_{j=0}^{\infty}$. The model to be solved is

$$a_t = \varphi\xi_t + \delta\int_0^1 \mathbb{E}^i\left[a_{t+1} \mid \Omega_t^i\right] di \tag{39}$$

$$\xi_t = B(L)\eta_t \tag{40}$$

where $\delta \equiv \beta+\gamma$, and where $\delta \equiv \beta+\gamma$ and $B(L) = B_0 + B_1 L + B_2 L^2 + \cdots$ is a square-summable lag polynomial, $\sum_j B_j^2 < \infty$. The innovation process $\{\eta_t\}$ is mean-zero white noise with finite variance σ_η^2. The more general specification of the exogenous process ξ_t is required for our results below.

Hierarchical Information. We assume a fraction $\mu \in (0,1)$ of *informed* agents who observe the primitive innovations directly, $\Omega^I = \{\eta_{t-j}\}_{j=0}^{\infty}$, and a complementary fraction $1-\mu$ of *uninformed* agents who observe only the history of the aggregate action, $\Omega_t^U = \{a_{t-j}\}_{j=0}^{\infty}$. The model is given by

$$a_t = \varphi\xi_t + \delta\mu\mathbb{E}^I\left[a_{t+1} \mid \{\eta_{t-j}\}_{j=0}^{\infty}\right] + \delta(1-\mu)\mathbb{E}^U\left[a_{t+1} \mid \{a_{t-j}\}_{j=0}^{\infty}\right] \tag{41}$$

with the exogenous process following (40).

Non-revelation. Due to the assumption of idiosyncratic shocks and a continuum of agents, the equilibrium of both informational specifications will exist in $a_t = A(L)\eta_t$ for some square-summable $A(L)$. Thus, the history of the endogenous variable a_t must not reveal the history of the fundamental innovation η_t. Let $\mathcal{H}_t^\eta \equiv \overline{\text{span}}\{\eta_{t-j}\}_{j\geq 0}$ and $\mathcal{H}_t^a \equiv \overline{\text{span}}\{a_{t-j}\}_{j\geq 0}$ denote the span of these spaces, the non-revelation condition is $\mathcal{H}_t^a \subsetneq \mathcal{H}_t^\eta$. Equivalently, the mapping from η_t to a_t must admit a *non-fundamental* moving-average representation. The minimal (scalar) specification consistent with this information structure is[3]

$$a_t = (L-\lambda)\widetilde{A}(L)\eta_t \qquad |\lambda| < 1,\ \lambda \neq \delta, \tag{42}$$

with $\widetilde{A}(L)$ one-sided, square-summable, and free of additional zeroes inside the unit circle. The restriction $\lambda \neq \delta$ rules out the degenerate case in which the informational root coincides with the forward-looking discount root.

The guess (42) is non-fundamental for η_t given that λ lies inside the unit circle. The Wold (fundamental) representation is obtained by flipping λ outside the unit circle through the use of a Blaschke factor $\mathscr{B}_\lambda(L) \equiv (L-\lambda)/(1-\lambda L)$

$$a_t = (L-\lambda)\widetilde{A}(L)\eta_t = (1-\lambda L)\widetilde{A}(L)\,e_t \tag{43}$$

$$e_t = \mathscr{B}_\lambda(L)\eta_t = (L-\lambda)\left(\eta_t + \lambda\eta_{t-1} + \lambda^2\eta_{t-2} + \cdots\right) \tag{44}$$

[3] There could be multiple λ's inside the unit circle but this only serves to complicate the algebra, which is why we focus on the scalar case.

Because $\mathscr{B}_\lambda(z)\mathscr{B}_\lambda(z^{-1}) = 1$ on the unit circle, the transformation leaves the spectral density of a_t unchanged. Observing the history of a_t is therefore informationally equivalent to observing the history of the Wold innovation e_t, and any projection onto $\overline{\text{span}}\{a_{t-j}\}_{j\geq 0}$ reduces to a projection onto $\overline{\text{span}}\{e_{t-j}\}_{j\geq 0} \subsetneq \mathscr{H}_t^\eta$. The moving-average root λ indexes the degree of information, as $|\lambda| \to 1$ the public history becomes fundamental and the economy collapses to full information, whereas for $|\lambda|$ less than unity the public signal a_t is strictly less informative than the underlying η_t, and each agent's private signal η_{it} retains incremental value. The value of λ is endogenously determined.[4]

Dispersed-Hierarchical Aggregate Equivalence. We now show that the dispersed information economy and hierarchical information economy are aggregate equivalent.

Proposition 2 (Endogenous-Signal Equilibrium)**.** *Consider either the dispersed-information economy* (39)–(40) *or the hierarchical-information economy* (41)–(40)*. Let* $\kappa \in (0,1)$ *denote the informational parameter where* $\kappa = \tau \equiv \sigma_\eta^2/(\sigma_\eta^2 + \sigma_v^2)$ *in the dispersed economy and* $\kappa = \mu$ *in the hierarchical economy. If* $\delta \in (0,1)$ *and there exists* $|\lambda| < 1$, $\lambda \neq \delta$, *such that*

$$B(\lambda) - \frac{\kappa\,\delta\,B(\delta)}{\kappa\lambda - (1-\kappa)\left(\dfrac{\delta-\lambda}{1-\lambda\delta}\right)} = 0 \tag{45}$$

then the rational-expectations equilibrium aggregate action is

$$a_t = \frac{\varphi}{L-\delta}\left\{LB(L) - \delta B(\delta)\left(\frac{\kappa\lambda - (1-\kappa)\mathscr{B}_\lambda(L)}{\kappa\lambda - (1-\kappa)\mathscr{B}_\lambda(\delta)}\right)\right\}\eta_t \tag{46}$$

with $\mathscr{B}_\lambda(L) \equiv \dfrac{L-\lambda}{1-\lambda L}$ *and* $\mathscr{B}_\lambda(\delta) \equiv \dfrac{\delta-\lambda}{1-\lambda\delta}$*. Consequently, the dispersed and hierarchical economies are aggregate-equivalent whenever* $\tau = \mu$*.*

Proof. See Appendix A.5.[5] □

Interpretation. For the uninformed to remain uninformed in equilibrium, *knowledge of the model* must not reveal the primitive shocks. This is the fixed point that distinguishes endogenous from exogenous information because agents can now learn from the equilibrium mapping. To see it concretely, note that an uninformed agent can subtract the term $\delta(1-\mu)\mathbb{E}_t^U[a_{t+1}]$ from the observed aggregate action

[4]The non-fundamental representation (42) is not a modeling shortcut. Rondina and Walker (2021) show that a pure non-fundamental process is informationally equivalent to an exogenous signal-plus-noise environment. The history of η_t itself acts as noise through *confounding dynamics*—an infinite past of primitive shocks confounds agents into persistent forecast errors even without a superimposed noise process. The guess (42) can be viewed as a reduced form that replicates the informational content of an economy in which agents observe fewer aggregate signals than underlying shocks.

[5]An immediate implication is that any purely autoregressive specification, $B(L) = 1/\Phi(L)$ for a polynomial Φ with all roots outside the unit circle, is purely revealing. Thus, the AR(1) process in Angeletos and Huo (2021) is fundamental for η_t and this applies to any autoregressive structure. However, there are multiple ways to preserve asymmetric information in equilibrium, such as adding measurement noise to the aggregate, observing fewer aggregates than shocks, or giving the fundamental a non-invertible moving-average component. (Rondina and Walker, 2021) demonstrate how these alternatives can be informationally equivalent.

in (41), yielding

$$a_t - \delta(1-\mu)\mathbb{E}_t^U[a_{t+1}] = \delta\mu\mathbb{E}_t^I[a_{t+1}] + \varphi\xi_t = \left\{\delta\mu L^{-1}\left[(L-\lambda)\widetilde{A}(L) + \lambda\widetilde{A}_0\right] + \varphi B(L)\right\}\eta_t \tag{47}$$

Equation (47) is the exact linear combination of structural shocks that uninformed agents can infer through endogenous signal extraction. For equilibrium to be consistent with heterogeneous information, this must convey no more than the history $e_t, e_{t-1}, \ldots$, which holds if and only if (47) vanishes at $L = \lambda$—verifying the guess $a_t = (L-\lambda)\widetilde{A}(L)\eta_t$. Condition (45) is exactly this requirement.

The representation in Theorem 2 is algebraically the cleanest because it makes the role of the information share μ transparent. Two limiting cases follow immediately: As $\mu \to 0$ the share of informed agents vanishes and the economy collapses to the homogeneous incomplete-information equilibrium and the exogenous process must satisfy $B(L) = (L-\lambda)\widetilde{B}(L)$ for some $|\lambda| < 1$, giving

$$a_t = \varphi\left(\frac{L(1-\lambda L)\widetilde{B}(L) - \delta(1-\lambda\delta)\widetilde{B}(\delta)}{L-\delta}\right)e_t \qquad e_t = \left(\frac{L-\lambda}{1-\lambda L}\right)\eta_t \tag{48}$$

As $\mu \to 1$, the equilibrium converges to the full-information rational-expectations solution,

$$a_t = \varphi\left(\frac{LB(L) - \delta B(\delta)}{L-\delta}\right)\eta_t \tag{49}$$

The full-information equilibrium (49) is the standard Hansen-Sargent (1980) formula. This equation displays the cross-equation restrictions known as the "hallmark" of rational expectations models, but there is also an informational interpretation to the H-S formula. The first component, $\varphi LB(L)/(L-\delta)$, is the perfect foresight equilibrium—assuming a representative agent, solve the model (41) forward, recursively substituting, imposing the law of iterated expectations and applying a no-bubble condition to arrive at

$$a_t = \mathbb{E}_t \sum_{j=0}^{\infty} \delta^j \varphi\xi_{t+j} = \mathbb{E}_t\left(\frac{\varphi LB(L)}{L-\delta}\right)\eta_t \tag{50}$$

If we appended the agent's information set with *future* values of η_t, such that agents have perfect foresight (PF) $\Omega_t^{PF} = \{\eta_{t-j}\}_{j=-\infty}^{\infty}$, (50) would be the rational expectations equilibrium. Therefore the last element of the H-S formula, $\varphi\delta B(\delta)/(L-\delta)$, represents the conditioning down associated with only observing current and past η_t's. Subtracting off this exact linear combination of future shocks, $\varphi\delta B(\delta)\sum_j \delta^j \eta_{t+j}$, stems from knowledge that the model is given by (50) and the information set of $\Omega_t^I = \{\eta_{t-j}\}_{j=0}^{\infty}$.[6] Thus, one can read off an information structure by studying the conditioning down component of the Hansen-Sargent formula. The next corollary takes advantage of this interpretation.

Corollary 2. *The equilibrium described in Theorem 2 has an equivalent representation in e space given by*

$$a_t = \frac{1}{L-\delta}\left\{(1-\lambda L)LH^U(L) - (1-\lambda\delta)\delta H^U(\delta)\right\}e_t \tag{51}$$

[6] As shown in Appendix A of Hansen and Sargent (1980), agents who know the model is given by (50) will form expectations optimally by subtracting off the principal part of the Laurent series expansion of $B(L)$ around δ, which is $\delta B(\delta)/(L-\delta)$.

where $H^U(L)e_t = (L-\lambda)^{-1}\{\varphi\xi_t - \mu\delta[a_{t+1} - \mathbb{E}_t^I(a_{t+1})]\}$. *And a representation in* η *space given by*

$$a_t = \frac{1}{L-\delta}\Big\{LH^I(L) - \delta H^I(\delta)\Big\}\eta_t \tag{52}$$

where $H^I(L)\eta_t = \varphi\xi_t - (1-\mu)\delta[a_{t+1} - \mathbb{E}_t^U(a_{t+1})]$.

Proof. Follows directly from Theorem 2. □

Innovations Representations and Speculative Dynamics. The two representations of Corollary 2 correspond to two economically distinct filtrations of the *same* observable a_t, one for informed and one for uninformed.[7] The informed agents' innovations representation, (52), lies in the η_t space

$$H^I(L)\eta_t = \varphi\xi_t - (1-\mu)\delta\big[a_{t+1} - \mathbb{E}_t^U(a_{t+1})\big] \tag{53}$$

and incorporates the uninformed agents' forecast error. Thus, even agents who see η_t directly must price in how the other type is mistaken, because those mistakes feed back into equilibrium through (41). In the uninformed agents' filtration, the innovation is e_t, not η_t, and the ARMA structure carries an extra $(1-\lambda L)$ MA factor that is absent from (52). The uninformed agents' fundamental is

$$H^U(L)\,e_t = (L-\lambda)^{-1}\big\{\varphi\xi_t - \mu\delta\big[a_{t+1} - \mathbb{E}_t^I(a_{t+1})\big]\big\} \tag{54}$$

with the $(L-\lambda)^{-1}$ factor encoding the endogenous signal-extraction problem. The uninformed cannot see η_t, but they can read off the informed agents' information once it is impounded into equilibrium prices and filtered through the Blaschke operator.

The upshot is that *speculative dynamics* appear as *fundamentals* in the Hansen–Sargent equation. One can recast the conditioning down interpretation as a function of *strategic interaction*. Obviously, such interpretation is not present in representative agent models. For example, standard asset-pricing theory identifies the fundamental as the discounted expected value of dividends; here, each agent type's fundamental incorporates a strategic component given by the other type's forecast.

Replicating the dispersed information structure. As in the static and dynamic exogenous-information economies, we can refine the coarse informational hierarchy while preserving the aggregate equilibrium exactly and approximating the dispersed cross section arbitrarily closely. The difference is that private information must now be constructed conditional on the information already revealed by the endogenous aggregate action following Proposition 2.

[7] A third, outside filtration—that of an econometrician who observes $\{a_{t-j}\}_{j\geq 0}$ but neither structural shocks nor type identifiers—coincides in closed linear span with the uninformed agents' filtration, $\mathscr{H}_t^{\text{econ}} = \mathscr{H}_t^U$. The econometrician's Wold decomposition yields $a_t = \widetilde{A}_0^{-1}[(1-\lambda L)\widetilde{A}(L)]\,w_t$ with $w_t = \widetilde{A}_0\,e_t$, an ARMA with the same roots as (51) but a different innovation variance and interpretation.

Theorem 4 (Endogenous Russian-Doll Representation)**.** *Let a_t^{D} denote the rational-expectations equilibrium of the dispersed-information economy characterized by Proposition 2 and define*

$$\tau \equiv \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2} \in (0,1)$$

and $\mathscr{C} \equiv \overline{\mathrm{span}}\{\eta_s : s \in \mathbb{Z}\}$. Fix any integer $J \geq 1$ and define the set of admissible hierarchical partitions

$$\Theta_J^{\mathrm{ES}}(\tau) \equiv \left\{(p,g) : p_j > 0, \quad \sum_{j=0}^{J} p_j = 1, \quad 0 < g_0 < \cdots < g_J < 1, \quad \sum_{j=0}^{J} p_j g_j = \tau\right\}$$

For any $(p,g) \in \Theta_J^{\mathrm{ES}}(\tau)$, define $\Pi_j \equiv \frac{g_j}{\sigma_\eta^2(1-g_j)}$ $\Delta_0 \equiv \Pi_0$, $\Delta_\ell \equiv \Pi_\ell - \Pi_{\ell-1} > 0$, $\ell = 1,\ldots,J$. Partition the population into sets $\mathscr{I}_j$, $j = 0,\ldots,J$, with shares p_j. For every agent i, construct the signal layers

$$s_{i,\ell,t} = \eta_t + \nu_{i,\ell,t}, \qquad \nu_{i,\ell,t} \sim \mathcal{N}\left(0, \Delta_\ell^{-1}\right), \qquad \ell = 0,\ldots,J$$

where the noises are independent across agents, layers, and dates and independent of η. Type j observes

$$\Omega_{i,j,t}^{\mathrm{H}} \equiv \sigma\left(a^{\mathrm{H},t}, s_{i,0}^t, \ldots, s_{i,j}^t\right), \qquad s_{i,\ell}^t \equiv \{s_{i,\ell,t}, s_{i,\ell,t-1}, \ldots\},$$

where $a^{\mathrm{H},t}$ is the history of the endogenous aggregate action in the hierarchical economy. These information sets are strictly nested: $\Omega_{i,0,t}^{\mathrm{H}} \subsetneq \Omega_{i,1,t}^{\mathrm{H}} \subsetneq \cdots \subsetneq \Omega_{i,J,t}^{\mathrm{H}}$.

Define the within-type and population-average expectations by

$$\overline{\mathbb{E}}_{j,t}^{\mathrm{H}}[Z] \equiv \frac{1}{p_j} \int_{\mathscr{I}_j} \mathbb{E}\left[Z \mid \Omega_{i,j,t}^{\mathrm{H}}\right] di, \qquad \overline{\mathbb{E}}_t^{\mathrm{H}}[Z] \equiv \sum_{j=0}^{J} p_j \overline{\mathbb{E}}_{j,t}^{\mathrm{H}}[Z],$$

and let $\mathbb{E}_t^I[Z] \equiv \mathbb{E}[Z \mid \eta^t]$ and $\mathbb{E}_t^U[Z] \equiv \mathbb{E}[Z \mid a^{\mathrm{D},t}]$. Then the following results hold.

(i) Exact average-expectation and higher-order-belief equivalence. *For every $Z \in \mathscr{C}$, every date t, and every type j,*

$$\overline{\mathbb{E}}_{j,t}^{\mathrm{H}}[Z] = g_j \mathbb{E}_t^I[Z] + (1-g_j)\mathbb{E}_t^U[Z], \qquad \overline{\mathbb{E}}_t^{\mathrm{H}}[Z] = \overline{\mathbb{E}}_t^{\mathrm{D}}[Z].$$

In particular,

$$\overline{\mathbb{E}}_t^{\mathrm{H}}[Z] = \tau \mathbb{E}_t^I[Z] + (1-\tau)\mathbb{E}_t^U[Z]$$

Consequently, the entire hierarchy of population-average fixed-date, time-shifted, and mixed-date higher-order beliefs coincides with its dispersed-information counterpart.

(ii) Exact aggregate equivalence. *For every $(p,g) \in \Theta_J^{\mathrm{ES}}(\tau)$, the rational-expectations aggregate equilibrium of the hierarchical economy satisfies $a_t^{\mathrm{H}} = a_t^{\mathrm{D}}$.*

(iii) Individual approximation. *Fix the shares p, let $\bar{j}_p \equiv \sum_{j=0}^{J} p_j j$, $g_j(\varepsilon) \equiv \tau + \varepsilon(j - \bar{j}_p)$, and define*

$$\bar{\varepsilon}_J \equiv \min\left\{\frac{\tau}{\bar{j}_p}, \frac{1-\tau}{J-\bar{j}_p}\right\} \tag{55}$$

For every $\varepsilon \in (0, \bar{\varepsilon}_J)$, $(p, g(\varepsilon)) \in \Theta_J^{\mathrm{ES}}(\tau)$. Let $a_{i,j,t}^{\mathrm{H}}(\varepsilon)$ denote the resulting equilibrium action of a type-j agent, let $a_{i,t}^{\mathrm{D}}$ denote an individual action in the dispersed economy, and define the aggregate-state filtration $\mathscr{F}_t^{\eta} \equiv \sigma(\eta_s : s \le t)$. As $\varepsilon \downarrow 0$,

$$\max_{0\le j\le J}\left|\mathbb{E}\left[a_{i,j,t}^{\mathrm{H}}(\varepsilon) \mid \mathscr{F}_t^{\eta}\right] - \mathbb{E}\left[a_{i,t}^{\mathrm{D}} \mid \mathscr{F}_t^{\eta}\right]\right| \longrightarrow 0 \tag{56}$$

$$\max_{0\le j\le J}\left|\mathrm{Var}\left(a_{i,j,t}^{\mathrm{H}}(\varepsilon) \mid \mathscr{F}_t^{\eta}\right) - \mathrm{Var}\left(a_{i,t}^{\mathrm{D}} \mid \mathscr{F}_t^{\eta}\right)\right| \longrightarrow 0 \tag{57}$$

The conditional-mean convergence holds almost surely and in L^2. Because the actions are conditionally Gaussian, their date-t conditional distributions converge weakly to the dispersed-information conditional distribution, uniformly across the finitely many types. These conditional Gaussian distributions can also be coupled to converge uniformly in conditional L^2.

Many of our previous results carry over with minor modifications; consider the identification result of Theorem 3. Conditional on the realized aggregate state, the only agent-specific component of a type-j action is the posterior $\hat{r}_{i,j,t}$, whose cross-sectional distribution is $\hat{r}_{i,j,t} \mid r_t, j \sim \mathcal{N}\left(g_j r_t, \sigma_r^2 g_j(1-g_j)\right)$. At a fixed date, individual behavior is therefore a finite Gaussian mixture in both locations and variances. To apply the earlier identification theorem, average also over the stationary distribution of the aggregate state. Because $r_t \sim \mathcal{N}(0, \sigma_r^2)$, and $\hat{r}_{i,j,t} \mid j \sim \mathcal{N}\left(0, \sigma_r^2 g_j\right)$ $\sigma_r^2 g_j = g_j^2\sigma_r^2 + \sigma_r^2 g_j(1-g_j)$. Thus, the variation in the component mean and the conditional variance combine to make the stationary type variance proportional to g_j, rather than to $g_j(1-g_j)$. In particular, remove the component common to all agents and define $Y_{i,j,t}^{\mathrm{ES}} \equiv a_{i,j,t}^{\mathrm{H}} - \varphi\xi_t - \delta\mathbb{E}_t^{U}\left[a_{t+1}^{\mathrm{D}}\right] = \delta\widetilde{A}_0\hat{r}_{i,j,t}$. Let $c_j^{\mathrm{ES}} \equiv g_j/\tau$. Since $\sum_j p_j g_j = \tau$, the relative gains satisfy $\sum_j p_j c_j^{\mathrm{ES}} = 1$, and

$$\frac{\mathbb{E}\left[\left(Y_{i,t}^{\mathrm{ES}}\right)^{2q}\right]}{(2q-1)!!\left(\delta^2\widetilde{A}_0^2\sigma_r^2\tau\right)^q} = \sum_{j=0}^{J} p_j\left(\frac{g_j}{\tau}\right)^q \tag{58}$$

This is exactly the moment problem in Theorem 3. Conditional on the benchmark gain τ, the stationary distribution of the private component of individual behavior therefore identifies the population shares and relative gains, and hence the gains themselves. Repeated fixed-date cross sections provide an additional identifying channel through the type-specific locations $g_j r_t$. And aggregate data remain uninformative about this partition. Every partition in $\Theta_J^{\mathrm{ES}}(\tau)$ generates the same aggregate action, the same public Wold innovation, and the same equilibrium condition (45).

3.2 IMPLICATIONS: INFORMATION TRANSMISSION Beyond tractability, the endogenous-signal environment of Section 3 allows us to study information transmission. In Morris and Shin (2002) and Angeletos and Huo (2021), the information structure is treated as a primitive, so these models do not ask how private information becomes impounded into an endogenous public signal. Here, that transmission is an equilibrium outcome. We derive the results using the informed–uninformed representation, which makes the mechanism transparent. Theorem 4 then extends the results to the general nested hierarchy.

Higher-Order Beliefs. To derive higher-order beliefs, iterate the equilibrium equation forward one period, $a_{t+1} = \delta\mu\mathbb{E}^I_{t+1}[a_{t+2}] + \delta(1-\mu)\mathbb{E}^U_{t+1}[a_{t+2}] + \varphi\xi_{t+1}$, and recall that the equilibrium takes the form $a_t = (L-\lambda)\widetilde{A}(L)\eta_t$. The appendix shows that the date-$(t+1)$ average expectation of the endogenous variable at $t+2$ equals its actual value minus the average market forecast error,

$$\mu\mathbb{E}^I_{t+1}a_{t+2} + (1-\mu)\mathbb{E}^U_{t+1}a_{t+2} = a_{t+2} + \mu\widetilde{A}_0\lambda\eta_{t+2} - (1-\mu)\widetilde{A}_0\mathscr{B}_\lambda(L)\eta_{t+2}. \tag{59}$$

The average error has two components. The first, $\widetilde{A}_0\lambda\eta_{t+2}$, is the error made by informed agents, weighted by their mass μ. The second, $\widetilde{A}_0\mathscr{B}_\lambda(L)\eta_{t+2} = \widetilde{A}_0 e_{t+2}$, is the error made by uninformed agents, weighted by $1-\mu$. While the form of the Blaschke factor shows that the uninformed agents' error contains innovations already known to the informed,

$$e_{t+2} = \mathscr{B}_\lambda(L)\eta_{t+2} = (L-\lambda)\big(\eta_{t+2} + \lambda\eta_{t+1} + \lambda^2\eta_t + \cdots\big),$$

in which only η_{t+1} and η_{t+2} are unknown to an informed agent at date t; every remaining term lies in Ω^I_t. Defining $r_t \equiv \eta_t - \mathbb{E}^U_t[\eta_t] = (1-\lambda^2)(1-\lambda L)^{-1}\eta_t$ it follows that $\mathbb{E}^I_t[e_{t+2}] = \lambda r_t$ and hence

$$\mathbb{E}^I_t\left(\overline{\mathbb{E}}_{t+1}a_{t+2}\right) = \mathbb{E}^I_t a_{t+2} - (1-\mu)\widetilde{A}_0\lambda r_t. \tag{60}$$

An informed agent's optimal forecast of next period's *consensus* therefore differs from their forecast of next period's *action*, because part of the average forecast error is predictable at date t. This is precisely why higher-order beliefs do not collapse into first-order beliefs. The uninformed, by contrast, cannot forecast either component of the average error: η_{t+2} is a future primitive innovation and e_{t+2} is a future Wold innovation in their own filtration. Their higher-order expectation therefore coincides with their first-order forecast, $\mathbb{E}^U_t\left(\overline{\mathbb{E}}_{t+1}a_{t+2}\right) = \mathbb{E}^U_t a_{t+2}$, even though they compute it correctly.

An immediate consequence is that the law of iterated expectations fails for the average expectation operator,

$$\overline{\mathbb{E}}_t\left(\overline{\mathbb{E}}_{t+1}a_{t+2}\right) = \overline{\mathbb{E}}_t a_{t+2} - \mu(1-\mu)\widetilde{A}_0\lambda r_t. \tag{61}$$

Iterating on this expression leads to the following proposition.

Proposition 3 (Higher-Order Average Expectations)**.** *Consider the endogenous hierarchical equilibrium of Theorem 4, expressed through the informed–uninformed representation of Proposition 2, and write $\widetilde{A}(L) = \sum_{k\geq 0}\widetilde{A}_k L^k$. For every integer $n \geq 1$,*

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\cdots\overline{\mathbb{E}}_{t+n}a_{t+n+1} = \overline{\mathbb{E}}_t[a_{t+n+1}] - (1-\mu)\left(\sum_{s=1}^{n}(\mu\lambda)^s\widetilde{A}_{n-s}\right)r_t \tag{62}$$

Proof. See Appendix A.7. □

Proposition 3 nests (61) as the case $n = 1$ and shows that higher-order beliefs remain analytically tractable in the hierarchical equilibrium, because the informed agents internalize the uninformed agents' forecast errors recursively. Each additional layer contributes a new response coefficient $\widetilde{A}_{n-s}$ and passes the inherited correction along at rate $\mu\lambda$, the product of the mass of informed agents and the degree of confounding in the public signal. Both are necessary: every term carries $(\mu\lambda)^s$ with $s \geq 1$, so the correction vanishes when either $\mu = 0$ or $\lambda = 0$. At $\lambda = 0$ the public innovation is simply the private innovation lagged one period, $e_t = \eta_{t-1}$ and $r_t = \eta_t$, so public information is stale rather than confounded, the uninformed agents' future error has no predictable tail, and nothing is transmitted through belief layers. We use (62) below to assess how much information is impounded into the equilibrium action through higher-order belief dynamics.

Information Revelation for an ARMA(1,1) Endogenous signal extraction plays a crucial role in models with heterogeneous beliefs but mechanisms of information transmission are typically intractable. Our analytical solutions permit analysis of information transmission which we exploit by calculating the exact informativeness of the signal needed to completely reveal the underlying state. That is, we can use the existence criteria of Theorem 2, specifically Equation (45), to determine the required τ or μ necessary to completely reveal the underlying shock sequence, η^t. Moreover, we can do so as a function of underlying parameters and as a function of higher-order beliefs. The change in this statistic with respect to these parameters and higher-order beliefs gives us an accurate measure of information transmission.

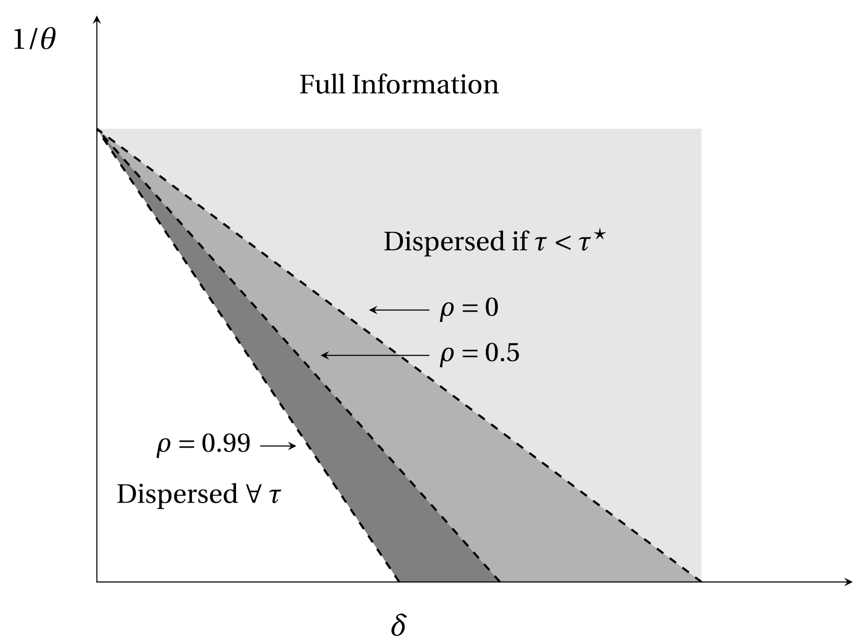

Figure 1: Existence space in $(\delta, 1/\theta)$. Existence of dispersed- and full-information equilibria following Corollary 3 for $\xi_t = \rho\xi_{t-1} + \eta_t + \theta\eta_{t-1}$. The dashed line is the exact boundary $1/\theta = 1 - \delta(1+\rho)$ of condition (64), plotted for $\rho \in \{0, 0.5, 0.99\}$; it is linear in these coordinates, which is why the vertical axis is drawn as $1/\theta$. Below and to the left of the line, dispersed information survives for every $\tau \in (0,1)$. Above and to the right, heterogeneity is preserved if and only if $\tau < \tau^\star$, with $\tau^\star$ given by (63). The region $1/\theta > 1$ corresponds to $\theta < 1$, where ξ_t is fundamental for η_t.

We begin with Figure 1, which characterizes the dispersed information equilibrium of Theorem 2 in the (δ, θ) space for the exogenous process, $\xi_t = \rho\xi_{t-1} + \eta_t + \theta\eta_{t-1}$. (Of course given our equivalence result, this figure also characterizes equilibrium for the hierarchical formulation of Theorem 2.) The figure is built around the following corollary to Theorem 2.

Corollary 3. *Consider the dispersed-information economy described by* (39) *of Proposition 2 with* $\xi_t = \rho\xi_{t-1} + \eta_t + \theta\eta_{t-1}$, $\delta, \rho \in (0,1)$ *and* $\theta > 0$*Define*

$$\tau^\star = \frac{(\theta - 1)(1 - \rho\delta)}{\delta(1+\rho)(1+\theta\delta)}. \tag{63}$$

The equilibrium is characterised in (δ, θ) *space as follows.*

(i) If $\theta \leq 1$*, then* $\tau^\star \leq 0$*: no dispersed-information equilibrium exists and the economy is characterised by the full-information equilibrium.*

(ii) If $\theta > 1$*, a dispersed-information equilibrium exists for every* $\tau \in (0,1)$ *if and only if*

$$1 - \delta(1+\rho) > 0 \qquad \text{and} \qquad \theta \geq \frac{1}{1 - \delta(1+\rho)}. \tag{64}$$

(iii) If $\theta > 1$ *and* (64) *fails, a dispersed-information equilibrium exists if and only if* $\tau \in (0, \tau^\star)$*, with* $\tau^\star \in (0,1)$.

Proof. See Appendix A.8. □

Three points are noteworthy. First, as is evident from Figure 1 and statement 1 of Corollary 3, if $\theta \leq 1$, the endogenous variable fully reveals the underlying shock, η_t, and the equilibrium is consistent with the full information equilibrium. With $\theta \leq 1$, confounding dynamics are not present in the exogenous process and ξ_t is fundamental for η_t. Second, from statement 3 and Figure 1, for a certain region of the parameter space (to the right of the dashed lines in Figure 1) a dispersed information equilibrium exists only if the signal-to-noise ratio is sufficiently small. The dashed lines represent the equilibrium that prevails as $\tau \to 1$, plotted for various serial correlation parameters. To the left of the dashed line, dispersed information will always be preserved in equilibrium regardless of the informativeness of the exogenous signal. The derivations of Section 3 demonstrate that an increase in θ may be interpreted as an increase in the noise associated with the endogenous signal extraction problem. The information content of the endogenous variable is sufficiently small that no matter how informative the exogenous signal, the full information equilibrium cannot be learned. How the discount factor δ alters the space of existence is similar to that of the serial correlation parameter ρ, which is the final point to be made. As the serial correlation in the ξ_t process increases and δ increases, it is more difficult to preserve dispersed information, *ceteris paribus* (the dashed line shifts to the left as ρ increases from 0 to 0.99). An increase in δ and ρ leads to a longer lasting effect of current information. This results in a higher $|\lambda|$ and a decrease in the informational discrepancy between fully informed and uninformed agent types.

HoBs and Information Transmission. Higher-order belief dynamics play a crucial role in disseminating information. Informed agents anticipate the component of the uninformed agents' forecast errors that is predictable from informed agents' information—the correction term in (60)—and act on it. Their date-t action therefore responds to information the uninformed do not have, the equilibrium action a_t inherits that response, and the uninformed subsequently observe a more informative public signal.

One consequence of this informational feedback effect is highlighted in Figure 2. This figure shows the existence space of the dispersed or hierarchical equilibria of Theorem 2 as higher-order belief dynamics are sequentially removed from the expectation of the informed agents. We modify the forecasting rule of the informed agents so that the law of iterated expectations is imposed at the first N nesting levels,

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\cdots\overline{\mathbb{E}}_{t+n}a_{t+n+1} \longmapsto \overline{\mathbb{E}}_t a_{t+n+1}, \qquad n = 1,\dots,N, \tag{65}$$

That is, we solve the equilibrium imposing that the law of iterated expectations holds at horizon $t = 1$, and derive the corresponding existence space given by Corollary 3. We then impose the law of iterated expectations at $t = 1,2$ and derive the existence space; impose the law of iterated expectations at horizons $t = 1,2,3$, and so forth. The x-axis indicates the horizons of HoBs removed. As HoBs are removed, the dispersed information equilibrium can support more informed agents or a higher signal-to-noise ratio. This is because the information that the uninformed are extracting from the endogenous variable is declining as fewer HoBs are being formulated. When we impose the law of iterated expectations on the entire dynamic structure (No HoBs or ∞ on the x-axis for Figure 2), the number of informed agents or the informativeness of the exogenous signal can nearly double (from 0.065 to 0.122) without fully revealing all underlying shocks. While the structure of the model will dictate the extent to which information can be impounded into equilibrium variables, endogenous signal extraction and the formation of higher-order beliefs plays an important role in information transmission.

Finally, although the preceding results were derived using the hierarchical representation, they apply directly to the symmetrically dispersed-information economy. The hierarchy in Theorem 4 is a device for solving the model: it replaces the difficult problem of agents forecasting the forecasts of others with an equivalent nested information structure. For every admissible hierarchy, the population-average expectation operator is the same as in the dispersed-information economy. Consequently, at every belief order,

$$\overline{\mathbb{E}}^{\mathrm{H}}_t\overline{\mathbb{E}}^{\mathrm{H}}_{t+1}\cdots\overline{\mathbb{E}}^{\mathrm{H}}_{t+n}a_{t+n+1} = \overline{\mathbb{E}}^{\mathrm{D}}_t\overline{\mathbb{E}}^{\mathrm{D}}_{t+1}\cdots\overline{\mathbb{E}}^{\mathrm{D}}_{t+n}a_{t+n+1}.$$

Proposition 3 therefore characterizes the higher-order beliefs of the dispersed-information economy itself, not merely those of an approximating hierarchy.

The relevant limit holds the number of information types fixed and lets the differences across their information intensities converge to zero. Along this limit, individual behavior converges to that of the symmetric dispersed-information economy, while the population-average expectation operator, the equilibrium law of motion $a_t = (L-\lambda)\widetilde{A}(L)\eta_t$, the root λ, and the revelation threshold $\mu^\star$ remain unchanged. Symmetrically dispersed information means that agents have the same signal precision; it does not mean that they observe the same signal. Agents continue to receive different private information and there-

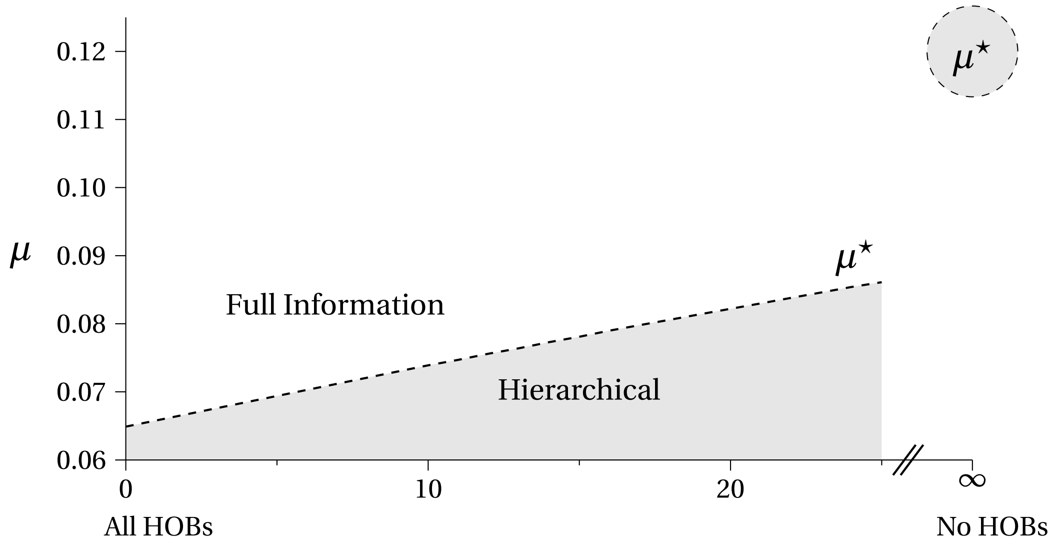

Figure 2: Full-revelation threshold under higher-order-belief truncation. The horizontal axis gives the number N of higher-order-belief layers suppressed. The case $N = 0$ is the rational-expectations equilibrium with the complete belief hierarchy, while $N \to \infty$ suppresses all higher-order-belief corrections. Parameters are $\xi_t = 0.8\xi_{t-1} + \eta_t + \sqrt{11}\eta_{t-1}$ and $\delta = 0.985$.

fore continue to forecast the beliefs and actions of others. What vanishes in the limit is the difference between the artificial types used to construct the hierarchy, not the higher-order-belief mechanism operating through dispersed private information.

4 SCOPE AND LIMITS OF THE EQUIVALENCE

Although our results are developed using the beauty-contest models of Morris and Shin (2002) and Angeletos and Huo (2021), the mechanism is more general and operational in many well-known environments. In this section, we discuss what assumptions and model structures are needed for our analysis to be applicable.

Assumption 1. Each agent's forecast of any common variable depends on that agent's information only through her own posterior mean, with weights that are the same for everyone. Formally, for every common variable Z entering the aggregate equilibrium and for each economy $e \in \{\mathrm{D}, \mathrm{H}\}$,

$$\mathbb{E}^{e}_{i,t}[Z] = \alpha_{Z,t} + \beta_{Z,t} m^{e}_{i,t} \tag{66}$$

where $\alpha_{Z,t}$ and $\beta_{Z,t}$ do not depend on i or on e, and the same representation holds when Z is replaced by any future population-average forecast of Z.

This assumption opens the door to the follow proposition.

Proposition 4 (Aggregate Compression)**.** *Let X_t denote the payoff-relevant state and let $\mathscr{B}_t$ denote common information, with $b_t \equiv \mathbb{E}[X_t \mid \mathscr{B}_t]$. Suppose that, in the dispersed economy,*

$$m^{\mathrm{D}}_{i,t} = (1 - g_i) b_t + g_i s_{i,t}, \qquad s_{i,t} = X_t + \nu_{i,t}, \qquad g_i \in [0,1], \tag{67}$$

and $\int_0^1 g_i \nu_{i,t}\, di = 0$. *Define* $\bar{g} \equiv \int_0^1 g_i\, di$.

(i) Contemporaneous compression. *The population-average posterior is*

$$\overline{\mathbb{E}}_t^{\mathrm{D}}[X_t] = (1-\bar{g})b_t + \bar{g}X_t. \tag{68}$$

This is also the average posterior in a two-tier hierarchy in which a fraction $\bar{g}$ *observes* X_t *and the remainder observes only* $\mathscr{B}_t$.

(ii) Dynamic equivalence. *If Assumption 1 holds and the aggregate equilibrium is linear in population-average expectations, then the two economies have the same population-average expectation hierarchy and the same aggregate equilibrium.*

Proof. Averaging gives

$$\begin{aligned}\int_0^1 m_{i,t}^{\mathrm{D}}\, di &= \int_0^1 (1-g_i)b_t\, di + \int_0^1 g_i X_t\, di + \int_0^1 g_i \nu_{i,t}\, di \\ &= (1-\bar{g})b_t + \bar{g}X_t.\end{aligned}$$

In the hierarchy, the uninformed posterior is b_t and the informed posterior is X_t, which proves part (i).

Under Assumption 1, for every relevant common variable Z, $\overline{\mathbb{E}}_t^{e}[Z] = \alpha_{Z,t} + \beta_{Z,t}\left[(1-\bar{g})b_t + \bar{g}X_t\right]$, with $e \in \{\mathrm{D},\mathrm{H}\}$. The same argument applies recursively to future average forecasts, so the population-average expectation hierarchies coincide at every order. Linearity of the aggregate equilibrium then proves part (ii). $\square$

These restrictions and assumptions hold in a broad set of models beyond the beauty-contest environments of Morris and Shin (2002) and Angeletos and Huo (2021) studied here. In macroeconomics, they are satisfied in the Lucas-islands tradition (Lucas, 1972), in models of rational inattention and limited information processing (Luo and Young, 2014; Mackowiak and Wiederholt, 2009; Gabaix and Laibson, 2022), in sticky-information models of price setting (Mankiw and Reis, 2002), and in models of nominal rigidity with dispersed information about aggregate demand (Angeletos and La'O, 2009). In finance, the same structure underlies the classical noisy rational-expectations literature on informationally efficient prices (Grossman and Stiglitz, 1980; Hellwig, 1980). In the analysis of policy and coordination, it appears in global games with endogenous information, where the informativeness of signals is itself determined in equilibrium (Angeletos et al., 2006). The empirical literature on information rigidities provides external discipline on the effective informed share (Coibion and Gorodnichenko, 2015), and recent survey evidence confirms that higher-order beliefs are distinct from first-order beliefs rather than redundant with them (Gorodnichenko and Yin, 2024). The last two cases require care, however: endogenous prices and endogenous information both make the informational anchor an equilibrium object, and we return to these below.

5 CONCLUDING COMMENTS

This paper shows that the aggregate content of dispersed information can be represented by a hierarchy of information. In the linear-Gaussian beauty-contest economies studied here, the cross-sectional distribution of beliefs enters aggregate dynamics only through a sufficient statistic and thus, the hierarchical economy preserves every order of average expectation and therefore the aggregate action. The Russian Doll refinement shows that nested information tiers preserve exact aggregate equivalence while approximating dispersed-information individual actions arbitrarily well. The equivalence is useful because it turns higher-order beliefs into tractable objects and allows for identification that is absent when agents are symmetrically informed. Extending this sufficient-statistic logic to multivariate DSGE, HANK, network, and endogenous information-acquisition environments is a natural next step.

REFERENCES

Acharya, Sushant, "Dispersed Beliefs and Aggregate Demand Management," February 2013.

__ **, Jess Benhabib, and Zhen Huo**, "The Anatomy of Sentiment-driven Fluctuations," *Journal of Economic Theory*, 2021, *195*, 105280.

Amador, Manuel and Pierre-Olivier Weill, "Learning from Prices: Public Communication and Welfare," *Journal of Political Economy*, 2010, *118* (5), 866–907.

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

__ **and Iván Werning**, "Crises and Prices: Information Aggregation, Multiplicity, and Volatility," *American Economic Review*, 2006, *96* (5), 1720–1736.

__ **and Jennifer La'O**, "Incomplete Information, Higher-Order Beliefs and Price Inertia," *Journal of Monetary Economics*, 2009, *56*, S19–S37. Supplement issue, October 2009.

__ **and** __ , "Sentiments," *Econometrica*, 2013, *81* (2), 739–780.

__ **and Zhen Huo**, "Myopia and Anchoring," *American Economic Review*, 2021, *111* (4), 1166–1200.

__ **, Christian Hellwig, and Alessandro Pavan**, "Signaling in a Global Game: Coordination and Policy Traps," *Journal of Political Economy*, 2006, *114* (3), 452–484.

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

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

Broer, Tobias and Alexandre N. Kohlhas, "Forecaster (Mis-)Behavior," *Review of Economics and Statistics*, 2024, *106* (5), 1334–1351.

Chahrour, Ryan and Kyle Jurado, "Revisiting the Forecasts of Others," *Journal of Monetary Economics*, November 2025, *155*, 103844. Supplemental issue: Informational Frictions in Macroeconomics (In honor of Robert E. Lucas, Jr.).

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

Futia, Carl A., "Rational Expectations in Stationary Linear Models," *Econometrica*, 1981, *49*, 171–192.

Gabaix, Xavier and David Laibson, "Myopia and Discounting," May 2022. NBER Working Paper 23254.

Gorodnichenko, Yuriy and Xiao Yin, "Higher-Order Beliefs and Risky Asset Holdings," Working Paper 32680, National Bureau of Economic Research July 2024.

Grossman, Sanford J. and Joseph E. Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 1980, *70*, 393–408.

Han, Zhao, Fei Tan, and Jieran Wu, "Analytic Policy Function Iteration," *Journal of Economic Theory*, 2022, *200*, 105395.

Hansen, Lars P. and Thomas J. Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 1980, *2*, 7–46.

Hellwig, Martin, "On the Aggregation of Information in Competitive Markets," *Journal of Economic Theory*, 1980, *22*, 477–498.

Huo, Zhen and Marcelo Pedroni, "A Single-Judge Solution to Beauty Contests," *American Economic Review*, 2020, *110*, 526–568.

— **and Naoki Takayama**, "Rational Expectations Models with Higher-Order Beliefs," *Review of Economic Studies*, 2025, *92* (5), 3138–3173.

Jurado, Kyle, "Rational Inattention in the Frequency Domain," *Journal of Economic Theory*, 2023, *208*, 105604.

Kasa, Kenneth, "Forecasting the Forecasts of Others in the Frequency Domain," *Review of Economic Dynamics*, 2000, *3*, 726–756.

— **, Todd B. Walker, and Charles H. Whiteman**, "Heterogeneous Beliefs and Tests of Present Value Models," *Review of Economic Studies*, 2014, *81*, 1137–1163.

Lorenzoni, Guido, "A Theory of Demand Shocks," *American Economic Review*, 2009, *99* (5), 2050–2084.

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

Luo, Yulei and Eric R. Young, "Signal Extraction and Rational Inattention," *Economic Inquiry*, 2014, *52* (2), 811–829.

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

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

Melosi, Leonardo, "Estimating Models with Dispersed Information," *American Economic Journal: Macroeconomics*, 2014, *6* (1), 1–31.

Miao, Jianjun, Jieran Wu, and Eric R. Young, "Macro-financial Volatility under Dispersed Information," *Theoretical Economics*, 2021, *16* (1), 275–315.

Morris, Stephen and Hyun Song Shin, "Social Value of Public Information," *American Economic Review*, 2002, *92* (5), 1521–1534.

Nimark, Kristoffer P., "Dynamic Higher Order Expectations," Technical Report 2017. CEPR Discussion Paper 11863.

Rondina, Giacomo, "Incomplete Information and Informational Equilibria," 2008.

— **and Todd B. Walker**, "Confounding Dynamics," *Journal of Economic Theory*, 2021, *196*, 1–33.

Tan, Fei and Todd B. Walker, "Solving Generalized Multivariate Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2015, *60*, 95–111.

Taub, Bart, "Aggregate Fluctuations as an Information Transmission Mechanism," *Journal of Economic Dynamics and Control*, 1989, *13* (1), 113–150.

Townsend, Robert M., "Forecasting the Forecasts of Others," *Journal of Political Economy*, 1983, *91*, 546–588.

Walker, Todd B., "How Equilibrium Prices Reveal Information in Time Series Models with Disparately Informed Traders," *Journal of Economic Dynamics and Control*, 2007, *31*, 512–537.

Woodford, Michael, "Imperfect Common Knowledge and the Effects of Monetary Policy," in Philippe Aghion, Roman Frydman, Joseph E. Stiglitz, and Michael Woodford, eds., *Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, Princeton, N.J.: Princeton University Press, 2003, pp. 26–58.

A PROOFS

A.1 PROOF OF PROPOSITION 1 The only remaining step is to show the aggregate action is identical, which is trivial given that the average expectation is the same across economies. The equilibrium admits the representation

$$\bar{a} = (1-r)\sum_{k=1}^{\infty} r^{k-1}\overline{\mathbb{E}}^k[\theta]$$

Substituting $\overline{\mathbb{E}}^k[\theta] = (1-\mu^k)y + \mu^k\theta$ gives

$$\bar{a} = (1-r)\sum_{k=1}^{\infty} r^{k-1}\left[(1-\mu^k)y + \mu^k\theta\right] = (1-r)\left[\sum_{k=1}^{\infty} r^{k-1}y - \sum_{k=1}^{\infty} r^{k-1}\mu^k y\right] + (1-r)\sum_{k=1}^{\infty} r^{k-1}\mu^k\theta$$

Using $\sum_{k=1}^{\infty} r^{k-1} = 1/(1-r)$ and $\sum_{k=1}^{\infty} r^{k-1}\mu^k = \mu/(1-r\mu)$, we obtain

$$\bar{a} = (1-r)\left[\frac{1}{1-r}y - \frac{\mu}{1-r\mu}y\right] + (1-r)\frac{\mu}{1-r\mu}\theta = y + \frac{\mu(1-r)}{1-r\mu}(\theta - y)$$

and when $\mu = \beta/(\alpha+\beta)$ we have

$$\bar{a} = \left(\frac{\beta(1-r)}{\alpha+\beta(1-r)}\right)\theta + \left(\frac{\alpha}{\alpha+\beta(1-r)}\right)y$$

Since the hierarchical economy reproduces the same hierarchy of higher-order average beliefs and hence the same aggregate action as the original Morris–Shin economy, the two economies are aggregate-equivalent, as claimed in Proposition 1.

A.2 PROOF OF COROLLARY 1 A partition has J free population shares, since the $J+1$ shares must sum to one, and J free information gains, since the $J+1$ gains must satisfy

$$\sum_{j=0}^{J} p_j g_j = \mu.$$

Hence $\Theta_J(\mu)$ has $2J$ degrees of freedom. The strict positivity and ordering restrictions are inequalities and therefore do not reduce its local dimension.

The moment-matching problem can be written in the familiar form

$$\mathscr{M}_J(\vartheta) - m = 0,$$

where $m \in \mathbb{R}^q$ is the target vector of cross-sectional moments. Thus, as in GMM, the partition parameters are chosen to satisfy a collection of moment conditions. The rank condition

$$\operatorname{rank} D\mathscr{M}_J(\bar{\vartheta}) = q$$

means that the q moments respond independently to the $2J$ free partition parameters near $\bar{\vartheta}$. Equivalently, one can select q partition parameters whose $q \times q$ Jacobian is nonsingular. Holding the remaining parameters fixed, the Inverse Function Theorem then implies that every moment vector sufficiently close to $\mathscr{M}_J(\bar{\vartheta})$ can be matched exactly.

Because this adjustment takes place within $\Theta_J(\mu)$, the aggregate-equivalence restriction remains satisfied. The hierarchy of average expectations and the aggregate equilibrium therefore remain exactly unchanged.

Finally, q moments cannot respond independently to only $2J$ free parameters unless

$$q \le 2J \qquad J \ge \left\lceil \frac{q}{2} \right\rceil$$

which completes the proof.

A.3 PROOF OF THEOREM 2 Under the normalization $\operatorname{Var}(\eta_t) = 1$, define[8]

$$\bar{c} \equiv \inf_{\omega\in[-\pi,\pi]} \frac{S_x(e^{i\omega})}{S_\xi(e^{i\omega})} = 1 + \sigma_u^2(1-\rho)^2$$

[8] At $c = \bar{c}$, the measurement-noise spectrum remains nonnegative but vanishes at frequency zero. Hence we impose $c < \bar{c}$ to maintain strict positivity.

This is precisely the upper bound ensuring that $S_x(z)/c - S_\xi(z)$ is nonnegative at every frequency; the bound is attained at frequency zero. Fix a $(p,c) \in \Theta_J^{\mathrm{AH}}$.

Step 1: Validity and strict nesting. Because $c_J < \bar{c}$,

$$S_{u,J}(z) = \frac{S_x(z)}{c_J} - S_\xi(z) > 0$$

at every frequency. Because $c_{\ell-1} < c_\ell$,

$$S_{\zeta,\ell}(z) = S_x(z)\left(\frac{1}{c_{\ell-1}} - \frac{1}{c_\ell}\right) > 0, \qquad \ell = 1,\dots,J.$$

The required stationary Gaussian noise processes therefore exist and are nondegenerate. Iterating the signal recursion gives

$$x_{i,j,t} = \xi_t + u_{i,J,t} + \sum_{\ell=j+1}^{J} \zeta_{i,\ell,t}.$$

Independence of the noise layers and telescoping yield

$$S_{u,j}(z) = \frac{S_x(z)}{c_j} - S_\xi(z),$$
$$S_{x,j}(z) = \frac{S_x(z)}{c_j}.$$

The signal recursion also implies

$$\sigma\big(x_{i,0}^t,\dots,x_{i,j}^t\big) = \sigma\big(x_{i,j}^t,\zeta_{i,1}^t,\dots,\zeta_{i,j}^t\big).$$

The revealed garbling increments are independent of $(\xi, x_{i,j})$, so

$$\mathbb{E}\big[Z \mid \Omega_{i,j,t}\big] = \mathbb{E}\big[Z \mid x_{i,j}^t\big] \qquad \text{for every } Z \in \mathscr{C}.$$

Because $x_{i,j-1} = x_{i,j} + \zeta_{i,j}$ with independent components, the residual spectrum of $\zeta_{i,j}$ given $x_{i,j-1}$ is strictly positive:

$$S_{\zeta_j \mid x_{j-1}}(z) = \frac{S_{\zeta,j}(z) S_{x,j}(z)}{S_{\zeta,j}(z) + S_{x,j}(z)} > 0.$$

Previous layers are independent of $(x_{i,j},\zeta_{i,j})$ and do not remove this residual uncertainty. Thus $\zeta_{i,j}^t$ is not measurable with respect to $\Omega_{i,j-1,t}$, establishing strict nesting.

Step 2: Average expectations and higher-order beliefs. Let $B(L)$ be the canonical factor of the AH signal spectrum and $w_{i,t} \equiv B(L)^{-1} x_{i,t}$. The sequence $\{w_{i,t-h}\}_{h\ge 0}$ is orthonormal and spans the same information space as x_i^t. Gaussianity therefore gives, for every $Z \in \mathscr{C}$,

$$\mathbb{E}[Z \mid x_i^t] = \sum_{h=0}^{\infty} \mathrm{Cov}(Z, w_{i,t-h}) w_{i,t-h}.$$

Define the common component

$$v_t \equiv \int_0^1 w_{i,t}\, di = B(L)^{-1}\xi_t.$$

Because the idiosyncratic component of $w_{i,t}$ is independent of every $Z \in \mathscr{C}$,

$$\mathrm{Cov}(Z, w_{i,t-h}) = \mathrm{Cov}(Z, v_{t-h}).$$

Averaging yields

$$\overline{\mathbb{E}}_t^{\mathrm{AH}}[Z] = \sum_{h=0}^{\infty} \mathrm{Cov}(Z, v_{t-h}) v_{t-h}.$$

For type j, $B_j(L) = B(L)/\sqrt{c_j}$. The corresponding innovations satisfy

$$w_{i,j,t} = \sqrt{c_j}\, B(L)^{-1} x_{i,j,t}, \qquad \frac{1}{p_j}\int_{\mathscr{I}_j} w_{i,j,t-h}\, di = \sqrt{c_j}\, v_{t-h}, \qquad \mathrm{Cov}(Z, w_{i,j,t-h}) = \sqrt{c_j}\, \mathrm{Cov}(Z, v_{t-h}).$$

Using sufficiency and averaging within type j therefore gives

$$\overline{\mathbb{E}}^{\mathrm{H}}_{j,t}[Z] = c_j \overline{\mathbb{E}}^{\mathrm{AH}}_t[Z].$$

Population weighting and $\sum_j p_j c_j = 1$ recover

$$\overline{\mathbb{E}}^{\mathrm{H}}_t[Z] = \overline{\mathbb{E}}^{\mathrm{AH}}_t[Z].$$

Every intermediate population-average expectation of a variable in $\mathscr{C}$ remains in $\mathscr{C}$. Iteration establishes equality of all population-average fixed-date, time-shifted, and mixed-date higher-order beliefs.

Step 3: Aggregate equilibrium. Let $\mathscr{B}$ be the space of bounded common square-integrable processes with norm $\|a\|_{\mathscr{B}} \equiv \sup_t \|a_t\|_2$. The hierarchical aggregate best response

$$\left(\mathscr{T}^{\mathrm{H}} a\right)_t \equiv \varphi \sum_{h=0}^{\infty} \beta^h \overline{\mathbb{E}}^{\mathrm{H}}_t[\xi_{t+h}] + \gamma \sum_{h=0}^{\infty} \beta^h \overline{\mathbb{E}}^{\mathrm{H}}_t[a_{t+h+1}]$$

is a contraction of modulus $\gamma/(1-\beta) < 1$. It therefore possesses a unique fixed point in $\mathscr{B}$. The AH aggregate belongs to $\mathscr{C}$ and, by the operator identity of Step 2, satisfies the hierarchical fixed-point equation. Uniqueness implies $a^{\mathrm{H}}_t = a^{\mathrm{AH}}_t$.

Step 4: Individual approximation. The approximation follows from $a_{i,j,t}(\varepsilon) = c_j(\varepsilon)\mathscr{A}(L)x_{i,j,t}$, where $\mathscr{A}(L)$ is the common AH individual-action filter. Conditional means therefore differ by $[c_j(\varepsilon)-1]\mathscr{A}(L)\xi_t$, while the conditional idiosyncratic action spectrum is

$$\left|\mathscr{A}(e^{-i\omega})\right|^2 \left[c_j S_x(e^{i\omega}) - c_j^2 S_\xi(e^{i\omega})\right]. \tag{69}$$

Both converge to their AH counterparts as $c_j \to 1$, uniformly across types because J is finite. More specifically, $c_{j+1}(\varepsilon) - c_j(\varepsilon) = \varepsilon > 0$ and $\sum_j p_j c_j(\varepsilon) = 1$. The bound $\varepsilon < \bar{\varepsilon}_J$ ensures $(p, c(\varepsilon)) \in \Theta^{\mathrm{AH}}_J$. Because the aggregates coincide, agents in both economies forecast the same common target

$$\chi_t \equiv \sum_{h=0}^{\infty} \beta^h \left(\varphi \xi_{t+h} + \gamma a^{\mathrm{AH}}_{t+h+1}\right) \in \mathscr{C}.$$

Let $K_a(L)$ be the stable causal Wiener–Kolmogorov filter of the AH economy, so $a^{\mathrm{AH}}_{i,t} = K_a(L)x_{i,t}$. For type j,

$$a_{i,j,t}(\varepsilon) = c_j(\varepsilon) K_a(L) x_{i,j,t}(\varepsilon).$$

Conditional on $\mathscr{F}^\xi_t$ the means are therefore

$$\mathbb{E}\left[a^{\mathrm{AH}}_{i,t} \mid \mathscr{F}^\xi_t\right] = K_a(L)\xi_t, \qquad \mathbb{E}\left[a_{i,j,t}(\varepsilon) \mid \mathscr{F}^\xi_t\right] = c_j(\varepsilon) K_a(L)\xi_t.$$

Uniform convergence $c_j(\varepsilon) \to 1$ yields the conditional-mean claim almost surely and in L^2. The idiosyncratic component of the type-j action has spectrum

$$S^{\mathrm{id}}_{a,j}(z;\varepsilon) = |K_a(z)|^2 \left[c_j(\varepsilon) S_x(z) - c_j(\varepsilon)^2 S_\xi(z)\right],$$

which converges in L^1 to $|K_a(z)|^2 \sigma_u^2$. Integrating establishes conditional-variance convergence. Gaussianity then implies weak convergence of the date-t conditional distributions, uniformly across types.

A.4 PROOF OF THEOREM 3 The proof proceeds in two steps. First, the moments of individual behavior reveal the moments of the cross-sectional distribution of information loadings. Second, a finite number of these loading moments recovers the loadings and population shares.

The proportional-spectrum construction gives

$$S_{x,j}(z) = \frac{S_x(z)}{c_j}.$$

Because $Y_{i,j,t} = c_j H(L) x_{i,j,t}$, the stationary variance of individual behavior within type j is

$$\begin{aligned}\mathrm{Var}(Y_{i,j,t} \mid j) &= \frac{c_j^2}{2\pi} \int_{-\pi}^{\pi} \left|H(e^{-i\omega})\right|^2 \frac{S_x(e^{i\omega})}{c_j}\, d\omega \\ &= c_j V_{\mathrm{AH}}.\end{aligned} \tag{70}$$

Thus, after averaging over the stationary aggregate history,

$$Y_{i,j,t} \mid j \sim \mathcal{N}\left(0, c_j V_{\mathrm{AH}}\right). \tag{71}$$

The stationary individual distribution is therefore a finite mixture of normal distributions whose variance multipliers are the information loadings c_j and whose probabilities are the population shares p_j.
For every integer $q \geq 1$, Gaussianity implies

$$\mathbb{E}\left[Y_{i,t}^{2q}\right] = (2q-1)!!V_{\mathrm{AH}}^{q} \sum_{j=0}^{J} p_j c_j^{q}. \tag{72}$$

The pooled variance is V_{AH} because $\sum_j p_j c_j = 1$. Consequently, the normalized even moments satisfy

$$m_0 \equiv 1, \qquad m_q \equiv \frac{\mathbb{E}[Y_{i,t}^{2q}]}{(2q-1)!!\,V_{\mathrm{AH}}^{q}} = \sum_{j=0}^{J} p_j c_j^{q}, \qquad q \geq 1, \tag{73}$$

with

$$m_1 = \sum_{j=0}^{J} p_j c_j = 1.$$

Thus, the normalized even moments of observed individual behavior are exactly the moments of the cross-sectional distribution of information loadings.
We first suppose that the number of types is known. Set $n \equiv J+1$ and define the polynomial whose roots are the type-specific loadings:

$$P(z) \equiv \prod_{j=0}^{J} (z - c_j) = z^n + b_{n-1} z^{n-1} + \cdots + b_1 z + b_0. \tag{74}$$

Because $P(c_j) = 0$ for every type,

$$0 = \sum_{j=0}^{J} p_j c_j^k P(c_j) = m_{k+n} + \sum_{\ell=0}^{n-1} b_\ell m_{k+\ell}, \qquad k = 0, \ldots, n-1. \tag{75}$$

These n equations form a linear system for the n unknown coefficients $b_0, \ldots, b_{n-1}$. Its coefficient matrix is

$$\boldsymbol{H}_n \equiv \begin{pmatrix} m_0 & m_1 & \cdots & m_{n-1} \\ m_1 & m_2 & \cdots & m_n \\ \vdots & \vdots & & \vdots \\ m_{n-1} & m_n & \cdots & m_{2n-2} \end{pmatrix}. \tag{76}$$

This matrix is nonsingular. To see why, consider any nonzero vector $v = (v_0, \ldots, v_{n-1})'$. Using (73),

$$\begin{aligned} v' \boldsymbol{H}_n v &= \sum_{j=0}^{J} p_j \left(v_0 + v_1 c_j + \cdots + v_{n-1} c_j^{n-1}\right)^2 \\ &> 0. \end{aligned} \tag{77}$$

The inequality follows because all shares are positive and a nonzero polynomial of degree at most $n-1$ cannot equal zero at the n distinct points $c_0, \ldots, c_J$. Hence (75) has a unique solution for $b_0, \ldots, b_{n-1}$. The roots of the resulting polynomial $P(z)$ recover the loadings $\{c_0, \ldots, c_J\}$, and the maintained ordering

$$0 < c_0 < \cdots < c_J$$

assigns the type labels uniquely.
Once the loadings are known, the shares solve

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ c_0 & c_1 & \cdots & c_J \\ c_0^2 & c_1^2 & \cdots & c_J^2 \\ \vdots & \vdots & & \vdots \\ c_0^J & c_1^J & \cdots & c_J^J \end{pmatrix} \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \\ p_J \end{pmatrix} = \begin{pmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \\ m_J \end{pmatrix}. \tag{78}$$

This system also has a unique solution because the loadings are distinct. Indeed, a failure of uniqueness would again produce a nonzero polynomial of degree at most J that vanishes at all $J+1$ loadings.
The polynomial system uses the moments $m_0, \ldots, m_{2n-1}$, where $2n-1 = 2J+1$. The first two moments, $m_0 = m_1 = 1$, are al-

ready known from the probability and aggregate-equivalence restrictions. Therefore, the $2J$ nontrivial moments $m_2, \ldots, m_{2J+1}$ recover the $J+1$ ordered loadings and their $J+1$ population shares uniquely. This proves part (i).

Now suppose that the number of types is unknown but finite. For each integer $r \geq 1$, construct the moment matrix

$$\boldsymbol{H}_r \equiv \left[m_{k+\ell}\right]_{k,\ell=0}^{r-1}. \tag{79}$$

This matrix can be written as

$$\boldsymbol{H}_r = \begin{pmatrix} 1 & \cdots & 1 \\ c_0 & \cdots & c_J \\ \vdots & & \vdots \\ c_0^{r-1} & \cdots & c_J^{r-1} \end{pmatrix} \begin{pmatrix} p_0 & & 0 \\ & \ddots & \\ 0 & & p_J \end{pmatrix} \begin{pmatrix} 1 & c_0 & \cdots & c_0^{r-1} \\ \vdots & \vdots & & \vdots \\ 1 & c_J & \cdots & c_J^{r-1} \end{pmatrix}. \tag{80}$$

Because the shares are positive and the loadings are distinct,

$$\operatorname{rank}(\boldsymbol{H}_r) = \min\{r, J+1\}. \tag{81}$$

In economic terms, the moment matrix continues to acquire an independent row and column until all distinct information types have been counted. After that point, additional rows contain no new independent information.

It follows that the number of types is identified by the first moment matrix that becomes singular:

$$J+1 = \min\{r \geq 1 : \det(\boldsymbol{H}_{r+1}) = 0\}. \tag{82}$$

For a hierarchy with $J+1$ types, this calculation uses moments through m_{2J+2}. Once J has been recovered, the polynomial system (75) recovers the loadings and (78) recovers the population shares. Hence the full moment sequence identifies (J, p, c) globally, proving part (ii).

For example, when $J = 1$ is known, m_2 and m_3 are sufficient. These are obtained from the fourth and sixth moments of stationary individual behavior, respectively. The fourth moment identifies the dispersion in information loadings, while the sixth moment distinguishes the population share attached to the high and low loadings. This completes the proof.

A.5 PROOF OF PROPOSITION 2

Write the equilibrium guess as $a_t = (L-\lambda)\widetilde{A}(L)\eta_t$. The conditional expectations for the informed and uninformed are given by

$$\begin{aligned} \mathbb{E}_t^I(a_{t+1}) &= L^{-1}\left[(L-\lambda)\widetilde{A}(L) + \lambda\widetilde{A}_0\right]\eta_t \\ \mathbb{E}_t^U(a_{t+1}) &= L^{-1}\left[(L-\lambda)\widetilde{A}(L) - \widetilde{A}_0\mathscr{B}_\lambda(L)\right]\eta_t \end{aligned}$$

Substituting the expectations into the equilibrium gives the z-transform in η_t space as

$$(z-\lambda)\widetilde{A}(z) = \delta\mu z^{-1}\left[(z-\lambda)\widetilde{A}(z) + \lambda\widetilde{A}_0\right] + \delta(1-\mu)z^{-1}\left[(z-\lambda)\widetilde{A}(z) - \widetilde{A}_0\mathscr{B}_\lambda(z)\right] + \varphi B(z)$$

and re-arranging yields the following functional equation

$$(z-\lambda)(z-\delta)\widetilde{A}(z) = z\varphi B(z) + \delta\widetilde{A}_0\left[\mu\lambda - (1-\mu)\mathscr{B}_\lambda(z)\right]$$

The $\widetilde{A}(\cdot)$ process will not be analytic unless the process vanishes at the poles $z = \{\lambda, \delta\}$. Evaluating at $z = \lambda$ gives the restriction on $B(\cdot)$,

$$\varphi B(\lambda) = -\delta\mu\widetilde{A}_0.$$

Rearranging terms

$$\begin{aligned} (z-\delta)\widetilde{A}(z) &= \frac{1}{z-\lambda}\left\{z\varphi B(z) + \delta\widetilde{A}_0\left[\mu\lambda - (1-\mu)\mathscr{B}_\lambda(z)\right]\right\} \\ &= \frac{1}{z-\lambda}\left\{z\varphi B(z) + \delta\widetilde{A}_0 h(z)\right\} \end{aligned} \tag{83}$$

where

$$h(z) \equiv \mu\lambda - (1-\mu)\mathscr{B}_\lambda(z).$$

Evaluating at $z = \delta$ gives

$$\widetilde{A}_0 = -\frac{\varphi B(\delta)}{h(\delta)}$$

to ensure stability. This implies that the restriction on $B(\cdot)$ is

$$B(\lambda) = \frac{\delta\mu B(\delta)}{h(\delta)}$$

which is (45). Substituting this into (83) delivers (46).

For the dispersed economy, the first step is to obtain an innovations representation for the signal vector (η_{it}, a_t), i.e. to identify the space spanned by $\{\eta_{i,t-j}, a_{t-j}\}_{j=0}^{\infty}$. In terms of the primitive innovation η_t and the idiosyncratic noise v_{it}, the mapping is

$$\begin{pmatrix}\eta_{it}\\ a_t\end{pmatrix} = \begin{pmatrix}\sigma_\eta & \sigma_v\\ (L-\lambda)\widetilde{A}(L) & 0\end{pmatrix}\begin{pmatrix}\hat{\eta}_t\\ \hat{v}_{it}\end{pmatrix} = \Gamma(L)\begin{pmatrix}\hat{\eta}_t\\ \hat{v}_{it}\end{pmatrix} \tag{84}$$

rescaled so that $\hat{\eta}_t$ and $\hat{v}_{it}$ have unit variance. Denote the fundamental (Wold) representation by

$$\begin{pmatrix}\eta_{it}\\ a_t\end{pmatrix} = \Gamma^*(L)\begin{pmatrix}w_{it}^1\\ w_{it}^2\end{pmatrix} \tag{85}$$

As in the hierarchical case, we use a Blaschke factor to flip the non-fundamental root λ outside the unit circle, combined with a Gram–Schmidt orthogonalization W_λ that prevents the Blaschke transformation from introducing additional unstable roots:

$$W_\lambda = \frac{1}{\sqrt{\sigma_\eta^2+\sigma_v^2}}\begin{pmatrix}\sigma_\eta & -\sigma_v\\ \sigma_v & \sigma_\eta\end{pmatrix}, \qquad \mathscr{B}_\lambda(L) = \begin{pmatrix}1 & 0\\ 0 & \dfrac{1-\lambda L}{L-\lambda}\end{pmatrix}, \qquad \Gamma^*(L) = \Gamma(L)W_\lambda\mathscr{B}_\lambda(L).$$

The fundamental innovations are

$$\begin{pmatrix}w_{it}^1\\ w_{it}^2\end{pmatrix} = \mathscr{B}_\lambda(L^{-1})W_\lambda^T\begin{pmatrix}\hat{\eta}_t\\ \hat{v}_{it}\end{pmatrix}.$$

Applying the Wiener–Kolmogorov prediction formula to (85),

$$\mathbb{E}[a_{t+1} \mid \eta_i^t, a^t] = \left[\Gamma_{21}^*(L) - \Gamma_{21}^*(0)\right]L^{-1}w_{it}^1 + \left[\Gamma_{22}^*(L) - \Gamma_{22}^*(0)\right]L^{-1}w_{it}^2, \tag{86}$$

with

$$\Gamma_{21}^*(L) = (L-\lambda)\widetilde{A}(L)\frac{\sigma_\eta}{\sqrt{\sigma_\eta^2+\sigma_v^2}}, \qquad \Gamma_{21}^*(0) = -\lambda\widetilde{A}_0\frac{\sigma_\eta}{\sqrt{\sigma_\eta^2+\sigma_v^2}},$$

$$\Gamma_{22}^*(L) = -(1-\lambda L)\widetilde{A}(L)\frac{\sigma_v}{\sqrt{\sigma_\eta^2+\sigma_v^2}}, \qquad \Gamma_{22}^*(0) = -\widetilde{A}_0\frac{\sigma_v}{\sqrt{\sigma_\eta^2+\sigma_v^2}}.$$

Averaging across agents, the idiosyncratic components vanish and two terms that depend only on the aggregate innovation remain:

$$\int_0^1 w_{it}^1\,di = w_t^1 = \frac{\sigma_\eta}{\sqrt{\sigma_\eta^2+\sigma_v^2}}\hat{\eta}_t, \qquad \int_0^1 w_{it}^2\,di = w_t^2 = -\frac{\sigma_v}{\sqrt{\sigma_\eta^2+\sigma_v^2}}\frac{L-\lambda}{1-\lambda L}\hat{\eta}_t.$$

The average expectation is therefore

$$\bar{\mathbb{E}}[a_{t+1}] = \left[(L-\lambda)\widetilde{A}(L) + \lambda\widetilde{A}_0\right]L^{-1}\frac{\sigma_\eta^2}{\sigma_\eta^2+\sigma_v^2}\hat{\eta}_t + \left[(1-\lambda L)\widetilde{A}(L) - \widetilde{A}_0\right]L^{-1}\frac{\sigma_v^2}{\sigma_\eta^2+\sigma_v^2}\frac{L-\lambda}{1-\lambda L}\hat{\eta}_t. \tag{87}$$

Let $\tau \equiv \sigma_\eta^2/(\sigma_\eta^2+\sigma_v^2)$ and substitute (87) into the aggregate Euler equation (39), using $\xi_t = B(L)\eta_t$:

$$(L-\lambda)\widetilde{A}(L) = \delta\tau L^{-1}\left[(L-\lambda)\widetilde{A}(L) + \lambda\widetilde{A}_0\right] + \delta(1-\tau)L^{-1}\left[(L-\lambda)\widetilde{A}(L) + \widetilde{A}_0\frac{\lambda-L}{1-\lambda L}\right] + \varphi B(L)\sigma_\eta.$$

Setting $\widetilde{A}(L) = \varphi Q(L)\sigma_\eta$, the z-transform of the fixed-point condition in η-space is

$$(z-\lambda)Q(z) = \delta\tau z^{-1}[(z-\lambda)Q(z) + \lambda Q_0] + \delta(1-\tau)z^{-1}\left[(z-\lambda)Q(z) + Q_0\frac{\lambda-z}{1-\lambda z}\right] + B(z). \tag{88}$$

Rearranging,

$$(z-\lambda)(z-\delta)Q(z) = zB(z) + \delta Q_0\left[\tau\lambda + (1-\tau)\frac{\lambda-z}{1-\lambda z}\right]. \tag{89}$$

Analyticity of $Q(\cdot)$ requires the right-hand side to vanish at the poles $z \in \{\lambda, \delta\}$. Evaluating at $z = \lambda$ gives $B(\lambda) = -\delta\tau Q_0$, i.e.

$Q_0 = -B(\lambda)/(\delta\tau)$ when combined with the $z = \delta$ condition below. Defining $h(z) \equiv \tau\lambda + (1-\tau)(\lambda - z)/(1-\lambda z)$ and rearranging,

$$(z-\delta)Q(z) = \frac{1}{z-\lambda}\left[zB(z) + \delta Q_0\, h(z)\right]. \tag{90}$$

Evaluating at $z = \delta$ yields $Q_0 = -B(\delta)/h(\delta)$, which ensures stability and uniqueness. The fixed point for λ then becomes

$$B(\lambda) = \frac{\delta\tau B(\delta)}{h(\delta)},$$

which is (45). Substituting $Q_0 = -B(\delta)/h(\delta)$ into (90) and using $\widetilde{A}(L) = \varphi Q(L)\sigma_\eta$ delivers (46), completing the proof.

A.6 PROOF OF THEOREM 4

Proof. We verify the hierarchical equilibrium by conjecturing $a_t^{\mathrm{H}} = a_t^{\mathrm{D}}$, so that the endogenous public history observed by hierarchical agents is $a^{\mathrm{D},t}$. Because $(1-\lambda L)\widetilde{A}(L)$ is causally invertible, $a^{\mathrm{D},t}$ and e^t generate the same information.

Step 1: the sufficient private signal. For a type-j agent the signal layers combine into the precision-weighted statistic $\bar{s}_{i,j,t} \equiv \Pi_j^{-1}\sum_{\ell=0}^{j}\Delta_\ell s_{i,\ell,t} = \eta_t + \bar{v}_{i,j,t}$ with $\mathrm{Var}(\bar{v}_{i,j,t}) = \Pi_j^{-1}$. The remaining independent contrasts among layers eliminate η_t and contain only idiosyncratic noise, so they carry no information about common variables.

Step 2: the component missing from the public history. Let $r_t \equiv \eta_t - \mathbb{E}_t^U[\eta_t]$. Since $e_t = (L-\lambda)(1-\lambda L)^{-1}\eta_t$ is the public Wold innovation, $\mathbb{E}_t^U[\eta_t] = -\lambda e_t$ and

$$r_t = \eta_t + \lambda e_t = \frac{1-\lambda^2}{1-\lambda L}\eta_t, \qquad \sigma_r^2 \equiv \mathrm{Var}(r_t) = \sigma_\eta^2(1-\lambda^2), \qquad \eta_{t-k} - \mathbb{E}_t^U[\eta_{t-k}] = \lambda^k r_t \tag{91}$$

for every $k \geq 0$. Conditional on the public history the entire unresolved primitive history is therefore the scalar r_t; when $\lambda = 0$ this reduces to $e_t = \eta_{t-1}$ and $r_t = \eta_t$, so public information is one period behind. Removing what the public history already reveals, the type-j private history becomes $\bar{s}_{i,j,t-k} - \mathbb{E}_t^U[\eta_{t-k}] = \lambda^k r_t + \bar{v}_{i,j,t-k}$, which supplies total precision $\Pi_j\sum_{k\geq 0}\lambda^{2k} = \Pi_j/(1-\lambda^2)$ about r_t. Combining this with $\sigma_r^2 = \sigma_\eta^2(1-\lambda^2)$ gives the posterior loading

$$\frac{\sigma_\eta^2(1-\lambda^2)\Pi_j/(1-\lambda^2)}{1+\sigma_\eta^2(1-\lambda^2)\Pi_j/(1-\lambda^2)} = \frac{\sigma_\eta^2\Pi_j}{1+\sigma_\eta^2\Pi_j} = g_j: \tag{92}$$

the reduction in prior variance caused by the endogenous public signal is exactly offset by the information in the infinite private history. Writing $\widehat{r}_{i,j,t} \equiv \mathbb{E}[r_t \mid \Omega_{i,j,t}^{\mathrm{H}}]$, Gaussian projection and the continuum law of large numbers give $p_j^{-1}\int_{\mathscr{I}_j}\widehat{r}_{i,j,t}\,di = g_j r_t$ and, conditional on $\mathscr{F}_t^\eta$, $\widehat{r}_{i,j,t} \sim \mathcal{N}(g_j r_t, \sigma_r^2 g_j(1-g_j))$. The variance here is the cross-sectional dispersion of posterior means given the realized aggregate state; an individual agent's posterior variance of r_t is instead $\sigma_r^2(1-g_j)$.

Step 3: average expectations. For every $Z \in \mathscr{C}$ the informed–uninformed forecast gap lies in the one-dimensional space spanned by r_t, so $\mathbb{E}_t^I[Z] - \mathbb{E}_t^U[Z] = q_{Z,t}r_t$ for some scalar $q_{Z,t}$, and a type-j agent forecasts $\mathbb{E}[Z \mid \Omega_{i,j,t}^{\mathrm{H}}] = \mathbb{E}_t^U[Z] + q_{Z,t}\widehat{r}_{i,j,t}$. Averaging within type gives

$$\overline{\mathbb{E}}_{j,t}^{\mathrm{H}}[Z] = \mathbb{E}_t^U[Z] + g_j q_{Z,t} r_t = g_j\mathbb{E}_t^I[Z] + (1-g_j)\mathbb{E}_t^U[Z]. \tag{93}$$

In the dispersed economy each agent observes $\eta_{it} = \eta_t + v_{it}$, with per-period precision σ_v^{-2} about η_t, together with its entire history. Applying Step 2 with $\Pi = \sigma_v^{-2}$ yields loading $\sigma_\eta^2\sigma_v^{-2}/(1+\sigma_\eta^2\sigma_v^{-2}) = \tau$, hence $\overline{\mathbb{E}}_t^{\mathrm{D}}[Z] = \tau\mathbb{E}_t^I[Z] + (1-\tau)\mathbb{E}_t^U[Z]$. Averaging (93) across types and using $\sum_j p_j g_j = \tau$ gives

$$\overline{\mathbb{E}}_t^{\mathrm{H}}[Z] = \overline{\mathbb{E}}_t^{\mathrm{D}}[Z]. \tag{94}$$

Step 4: the aggregate equilibrium. Because $a_{t+1}^{\mathrm{D}} \in \mathscr{C}$, (94) gives

$$\begin{aligned}\varphi\xi_t + \delta\overline{\mathbb{E}}_t^{\mathrm{H}}\left[a_{t+1}^{\mathrm{D}}\right] &= \varphi\xi_t + \delta\overline{\mathbb{E}}_t^{\mathrm{D}}\left[a_{t+1}^{\mathrm{D}}\right]\\ &= a_t^{\mathrm{D}}.\end{aligned}$$

Thus, a_t^{D} satisfies the hierarchical equilibrium condition, and the conjectured public history $a^{\mathrm{D},t}$ is the public history it generates. The conjecture is therefore self-consistent: the hierarchical economy admits

$$a_t^{\mathrm{H}} = a_t^{\mathrm{D}}$$

as an equilibrium, with the same informational root λ and hence the same public Wold innovation e_t. Both economies are solved at the root of (45) fixed in the hypothesis of the theorem, so the equivalence is between corresponding equilibria rather than between arbitrary selections. This proves part (ii) and establishes (**??**) in equilibrium.

Each population-average expectation mapping sends $\mathscr{C}$ into itself. Repeated application of (94) therefore proves by induction that the two economies share all population-average fixed-date, time-shifted, and mixed-date higher-order beliefs. This proves part (i).

Step 5: individual approximation. Under (**??**), $\sum_j p_j g_j(\varepsilon) = \tau + \varepsilon \sum_j p_j (j - \bar{j}_p) = \tau$, and (55) ensures $0 < g_0(\varepsilon) < \cdots < g_J(\varepsilon) < 1$, so $(p, g(\varepsilon)) \in \Theta_J^{\mathrm{ES}}(\tau)$. Write $(L-\lambda)\widetilde{A}(L) = \sum_{k\geq 0} A_k L^k$ and $\widetilde{A}(L) = \sum_{k\geq 0} \widetilde{A}_k L^k$, so that $A_{k+1} = \widetilde{A}_k - \lambda \widetilde{A}_{k+1}$ and the sum telescopes,

$$\sum_{k=0}^{\infty} A_{k+1}\lambda^k = \sum_{k=0}^{\infty}\left(\widetilde{A}_k\lambda^k - \widetilde{A}_{k+1}\lambda^{k+1}\right) = \widetilde{A}_0, \tag{95}$$

where the last equality uses $|\lambda| < 1$ and square-summability of $\widetilde{A}(L)$, and holds trivially at $\lambda = 0$. With (91) this gives $\mathbb{E}_t^I[a_{t+1}^{\mathrm{D}}] - \mathbb{E}_t^U[a_{t+1}^{\mathrm{D}}] = \left(\sum_k A_{k+1}\lambda^k\right) r_t = \widetilde{A}_0 r_t$, so the type-$j$ action is $a_{i,j,t}^{\mathrm{H}}(\varepsilon) = \varphi\xi_t + \delta\mathbb{E}_t^U[a_{t+1}^{\mathrm{D}}] + \delta\widetilde{A}_0 \hat{r}_{i,j,t}(\varepsilon)$. Conditional on $\mathscr{F}_t^{\eta}$ it is Gaussian with

$$m_j(\varepsilon) = \varphi\xi_t + \delta\mathbb{E}_t^U[a_{t+1}^{\mathrm{D}}] + \delta\widetilde{A}_0 g_j(\varepsilon) r_t, \qquad v_j(\varepsilon) = \delta^2\widetilde{A}_0^2\sigma_r^2 g_j(\varepsilon)\big(1 - g_j(\varepsilon)\big), \tag{96}$$

and the dispersed action has the same expressions with $g_j(\varepsilon)$ replaced by τ. Since $\max_j |g_j(\varepsilon) - \tau| \to 0$, both converge uniformly across the finitely many types; the mean gap is proportional to $(g_j(\varepsilon) - \tau) r_t$, so its convergence holds almost surely and in L^2, proving (56)–(57). Conditional Gaussianity then delivers uniform weak convergence. For the coupling, let $z_i \sim \mathcal{N}(0,1)$ be independent of $\mathscr{F}_t^{\eta}$ and set $\tilde{a}_{i,j,t}^{\mathrm{H}}(\varepsilon) = m_j(\varepsilon) + \sqrt{v_j(\varepsilon)}\, z_i$ and $\tilde{a}_{i,t}^{\mathrm{D}} = m^{\mathrm{D}} + \sqrt{v^{\mathrm{D}}}\, z_i$; then

$$\mathbb{E}\left[\left(\tilde{a}_{i,j,t}^{\mathrm{H}}(\varepsilon) - \tilde{a}_{i,t}^{\mathrm{D}}\right)^2 \middle| \mathscr{F}_t^{\eta}\right] = \delta^2\widetilde{A}_0^2\left[\big(g_j(\varepsilon) - \tau\big)^2 r_t^2 + \sigma_r^2\left\{\sqrt{g_j(\varepsilon)\big(1 - g_j(\varepsilon)\big)} - \sqrt{\tau(1-\tau)}\right\}^2\right], \tag{97}$$

which converges to zero uniformly across types, almost surely for every fixed t. □

A.7 PROOF OF PROPOSITION 3

Proof. Write $\widetilde{A}(L) = \sum_{k=0}^{\infty}\widetilde{A}_k L^k$ and define

$$G_\lambda(L) \equiv \frac{1-\lambda^2}{1-\lambda L}.$$

Recall that $e_t = \mathscr{B}_\lambda(L)\eta_t$ and

$$a_t = (L-\lambda)\widetilde{A}(L)\eta_t = (1-\lambda L)\widetilde{A}(L)e_t.$$

Let

$$(1-\lambda L)\widetilde{A}(L) = \sum_{k=0}^{\infty} d_k L^k, \qquad d_0 = \widetilde{A}_0, \qquad d_k = \widetilde{A}_k - \lambda\widetilde{A}_{k-1} \quad \text{for } k \geq 1.$$

For every $s \geq 1$, the informed agents can forecast the component of the future public innovation inherited from past structural innovations:

$$\mathbb{E}_t^I[e_{t+s}] = \lambda^{s-1} G_\lambda(L)\eta_t. \tag{98}$$

By contrast, e_{t+s} is a future Wold innovation for the uninformed, so

$$\mathbb{E}_t^U[e_{t+s}] = 0. \tag{99}$$

Consider a target a_{t+h+1} with $h \geq 1$. At time $t+1$, the uninformed forecast error is

$$a_{t+h+1} - \mathbb{E}_{t+1}^U[a_{t+h+1}] = \sum_{k=0}^{h-1} d_k e_{t+h+1-k}.$$

Taking the informed agents' time-t expectation and using (98) gives

$$\begin{aligned}\mathbb{E}_t^I\left[\mathbb{E}_{t+1}^U[a_{t+h+1}]\right] &= \mathbb{E}_t^I[a_{t+h+1}] - G_\lambda(L)\eta_t\sum_{k=0}^{h-1} d_k\lambda^{h-k} \\ &= \mathbb{E}_t^I[a_{t+h+1}] - \lambda\widetilde{A}_{h-1}G_\lambda(L)\eta_t,\end{aligned} \tag{100}$$

where the second equality follows from

$$\sum_{k=0}^{h-1} d_k\lambda^{h-k} = \lambda\widetilde{A}_{h-1}.$$

The informed agents' own expectations satisfy the law of iterated expectations. It follows that

$$\mathbb{E}_t^I\left[\overline{\mathbb{E}}_{t+1} a_{t+h+1}\right] = \mathbb{E}_t^I[a_{t+h+1}] - (1-\mu)\lambda\widetilde{A}_{h-1}G_\lambda(L)\eta_t. \tag{101}$$

Because the uninformed cannot forecast either group's future forecast errors,

$$\mathbb{E}_t^U\left[\overline{\mathbb{E}}_{t+1}a_{t+h+1}\right] = \mathbb{E}_t^U[a_{t+h+1}]. \tag{102}$$

Averaging (101) and (102) therefore yields

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}a_{t+h+1} = \overline{\mathbb{E}}_t a_{t+h+1} - \mu(1-\mu)\lambda\widetilde{A}_{h-1}G_\lambda(L)\eta_t. \tag{103}$$

We also have

$$\overline{\mathbb{E}}_t\left[G_\lambda(L)\eta_{t+1}\right] = \mu\lambda G_\lambda(L)\eta_t. \tag{104}$$

Indeed, the informed expectation equals $\lambda G_\lambda(L)\eta_t$, whereas the uninformed expectation is zero because

$$\frac{1}{L-\lambda}e_{t+1} = \sum_{k=0}^{\infty}\lambda^k e_{t+2+k} \tag{105}$$

contains only future public innovations.

We now proceed by induction. For $j=1$, setting $h=1$ in (103) gives

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}a_{t+2} = \overline{\mathbb{E}}_t a_{t+2} - (1-\mu)(\mu\lambda)\widetilde{A}_0 G_\lambda(L)\eta_t,$$

which is the desired result.

Suppose the result holds at order $j-1$. Applying $\overline{\mathbb{E}}_t$ to that expression, using (103) with $h=j$ for its first term and (104) for every inherited correction, gives

$$\begin{aligned}
&\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\cdots\overline{\mathbb{E}}_{t+j}a_{t+j+1}\\
&\quad= \overline{\mathbb{E}}_t a_{t+j+1} - (1-\mu)\left[(\mu\lambda)\widetilde{A}_{j-1} + \sum_{i=1}^{j-1}(\mu\lambda)^{i+1}\widetilde{A}_{j-1-i}\right]G_\lambda(L)\eta_t\\
&\quad= \overline{\mathbb{E}}_t a_{t+j+1} - (1-\mu)\left(\sum_{i=1}^{j}(\mu\lambda)^i\widetilde{A}_{j-i}\right)G_\lambda(L)\eta_t.
\end{aligned}$$

This proves the result for every $j \geq 1$. □

A.8 HOBS COROLLARY

Substituting

$$B(z) = \frac{1+\theta z}{1-\rho z}$$

into the existence restriction (45) yields a cubic in λ whose coefficients depend on $(\delta,\rho,\theta,\tau)$. Two of its roots are inadmissible for the dispersed equilibrium: $\lambda=\delta$ is a removable root inherited from the $B(\delta)$ term, and the third root exceeds unity in modulus for all admissible parameters. The candidate non-fundamental root is therefore the remaining negative root, which is real and strictly decreasing in τ on $(0,1)$. Consequently the admissible region in τ is an interval containing zero, and its upper endpoint is the value of τ at which the root reaches the unit circle.

Because the root is negative and decreasing, it can only leave the unit disk through $\lambda=-1$. Evaluating (45) at $\lambda=+1$ and solving for τ gives

$$\tau\big|_{\lambda=1} = \frac{\theta+1-\delta\rho(\theta+1)}{\delta\big(1-\rho+\delta\theta-\delta\rho\theta\big)} > 1$$

for all $\delta,\rho\in(0,1)$ and $\theta>0$, so exit at $+1$ never occurs for admissible signal-to-noise ratios and no second case arises. Setting $\lambda=-1$ in (45) and solving for τ delivers (63). Hence a dispersed-information equilibrium exists if and only if $\tau\in(0,\tau^\star)$, intersected with $(0,1)$.

Part (i). The factors $1-\rho\delta$, $\delta(1+\rho)$ and $1+\theta\delta$ in (63) are strictly positive, so $\operatorname{sign}\tau^\star = \operatorname{sign}(\theta-1)$. If $\theta\leq 1$ then $\tau^\star\leq 0$ and the interval $(0,\tau^\star)$ is empty, leaving the full-information equilibrium as the unique equilibrium. The case $\theta=1$ is the knife edge at which $B(z)$ has a root exactly on the unit circle, $\tau^\star=0$; the moving-average component is then not invertible but neither is it confounding, and the fundamental representation obtains.

Part (ii). If $\theta>1$ then $\tau^\star>0$, and existence for every $\tau\in(0,1)$ holds if and only if $\tau^\star\geq 1$. Clearing the positive denominator of (63),

$$\tau^\star - 1 = \frac{(\theta-1)(1-\rho\delta)-\delta(1+\rho)(1+\theta\delta)}{\delta(1+\rho)(1+\theta\delta)},$$

and expanding the numerator gives

$$(\theta-1)(1-\rho\delta)-\delta(1+\rho)(1+\theta\delta) = (1+\delta)\Big(\theta\big[1-\delta(1+\rho)\big]-1\Big).$$

Since $1+\delta > 0$ and the denominator is positive, $\tau^\star \geq 1$ if and only if $\theta\left[1-\delta(1+\rho)\right] \geq 1$. This requires $1-\delta(1+\rho) > 0$, in which case it is equivalent to $\theta \geq \left[1-\delta(1+\rho)\right]^{-1}$, which is (64).

Part (iii). If $\theta > 1$ and (64) fails, the identity above gives $\tau^\star < 1$, so $\tau^\star \in (0,1)$ and existence holds if and only if $\tau \in (0,\tau^\star)$. □