# DISPERSED AND HIERARCHICAL ECONOMIES WITH DYNAMIC SIGNAL EXTRACTION: AN EQUIVALENCE RESULT*

Giacomo Rondina† Todd B. Walker‡

November 2023

**Abstract**

We derive an equivalence, in the aggregate, between dynamic models with dispersed and hierarchical information. Optimal signal extraction, in the dispersed case, suggests agents treat the signal as true with probability equal to the signal-to-noise ratio, and false with the complementary probability. Equivalence follows when the share of informed agents, in the hierarchical model, is set equal to the signal-to-noise ratio in the dispersed economy. The value of this theorem is due to the hierarchical model being easier to solve and interpret, especially when agents infer information from endogenous sources. We use our results to study the behavior of higher-order beliefs and information transmission in closed form in models with dispersed information and endogenous signal extraction. While we work within a generic environment, we show how our results can map into a well-known asset pricing model.

*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 comments on an earlier draft.

†Department of Economics, UCSD, grondina@ucsd.edu.

‡Department of Economics, Indiana University, walkertb@indiana.edu

# 1 INTRODUCTION

Several foundational papers in economics[1] emphasize the importance of informational frictions similar to the following problem: Consider the dynamic signal extraction problem of inferring a signal ($\varepsilon_t$) with superimposed noise ($\eta_t$) corrupting the signal. One observes the linear combination

$$\mathscr{S}_t = \varepsilon_t + \eta_t \tag{1}$$

and forms the conditional expectation, $\mathbb{E}(\varepsilon_t|\mathscr{S}^t)$, where $\mathscr{S}^t \equiv \mathscr{S}_t, \mathscr{S}_{t-1}, \ldots$. Assuming the signal and noise are drawn from mean-zero Gaussian distributions and are independent across time and uncorrelated at all leads and lags, then optimal signal extraction is well known with conditional expectation $\mathbb{E}(\varepsilon_t|\mathscr{S}^t) = \tau\mathscr{S}_t$, where $\tau = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$ is the relative weight given to the observed variable. An interpretation of this equation is that the agent employs a mixed strategy when forming expectations. She believes the observed variable is the true signal with probability given by the ratio of the variance of the signal to the signal plus noise variance, $\sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$. Alternatively, the agent believes the observed variable is pure noise in proportion $1-\tau$; that is, $\mathbb{E}(\eta_t|\mathscr{S}^t) = (1-\tau)\mathscr{S}_t = \sigma_\eta^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$. Thus the agent's behavior can be decomposed into disjointed actions: with probability $\tau$, the agent acts as perfectly informed and with probability $1-\tau$, the agent acts as perfectly uninformed. The clean delineation is due to the standard assumption that noise and signal are uncorrelated.

In many dynamic models with dispersed information, the source of the heterogeneity is generated by a continuum of agents in which each agent $i \in (0,1)$ observes some form of a noisy signal, $\mathscr{S}_{i,t} = \varepsilon_t + \eta_{i,t}$, where the noise is uncorrelated with the signal and the variance of the noise is identical across agents. In linear environments, endogenous variables are linear combinations of underlying shocks, which serves to operationalize an interpretation consistent with the disjointed behavior of optimal signal extraction described in the previous paragraph. The purpose of this paper is to show that models with dispersed information (i.e., models with a continuum of heterogeneous agents that are equally uninformed) can be mapped into models with hierarchical information structures (i.e., models were agents can be explicitly ranked according to the amount of information they possess). We refer to this result as the **Dispersed-Hierarchical Equivalence.** To the best of our knowledge, this paper is the first to prove such a connection.

Theorem 1 provides the equivalence, which holds in the aggregate, between models with dispersed information and models with hierarchical information. In the dispersed environment, there is a continuum of agents with each receiving an idiosyncratic noisy signal about the underlying state, coupled with information gleaned from endogenous sources. In the hierarchical setup, there are two types of agents: perfectly informed and uninformed. The uninformed agents can only perform endogenous signal extraction and remain uninformed in equilibrium. Theorem 1 shows that the aggregate representations of these equilibria can be equated once the parameter measuring the proportion of agents perfectly informed in the hierarchical model is reinterpreted as the signal-to-noise ratio of the privately observed

[1] For example, see Muth (1960) on rational expectations; Lucas (1972) on monetary theory; Kydland and Prescott (1982) on business cycle analysis.

signal in the dispersed information economy. While individual forecasts maintain a well defined cross-sectional distribution of beliefs in the dispersed economy (Proposition 3), the idiosyncratic noise component does not survive aggregation, which delivers our aggregate equivalence.

Section 3 studies the implications of our equivalence result. Given that information can be ordered in models with hierarchical information, sufficient statistics are available and equilibria relatively straightforward to compute. Conversely, with dispersed information, there is no sense in which the state can be summarized compactly from the viewpoint of each individual agent. Thus, there is a mapping between the equilibria that can be exploited to better understand aggregate dynamics in dispersed-information economies. Equipped with an analytical characterization of the market equilibria under dispersed information due to Theorem 1, we are able to characterize the higher-order belief (HoBs) representation of such equilibria in closed form and study the role of HoB thinking in the transmission of information. From the viewpoint of an arbitrary agent $i$, the optimality of signal extraction behooves her to act as informed with probability equal to the signal-to-noise ratio. In so doing, she will recognize that a fraction of agents is contemporaneously acting as uninformed. It follows that as an informed agent, she should forecast the forecast error of the agents acting as uninformed and embed it into her expectations about the future. She will adjust her time-$t$ forecast according to the collective ignorance of the uninformed agents, despite the fact that she is contributing to this collective ignorance. She correctly views her individual forecast error as infinitesimal in this regard and thus irrelevant for her reasoning.

We also use our closed-form solutions to study information transmission by calculating the informativeness of the exogenous signal just necessary to perfectly reveal the underlying state (an alternative interpretation, due to Theorem 1, is the exact fraction of informed agents necessary for perfect revelation of the underlying state). We show how to solve for this statistic as a function of model parameters, and then examine how this statistic changes with respect to model characteristics (Corollary 3). An increase in the discount factor or in the autocorrelation of the exogenous shock substantially facilitates information transmission. Because agents are learning endogenously from the forecast errors of other agent types, an increase in the persistence of these errors improves learning. Increasing the discount factor and autocorrelation parameter promotes this persistence in errors.

To understand the extent to which higher-order beliefs (HoBs) play a role in information dissemination, we sequentially remove HoBs from the model and calculate our statistic of information transmission. We first remove HoBs of order one only (i.e, informed agents time-$t$ expectation of the uninformed's $t+1$ forecast error) and calculate the share of informed agents necessary to fully reveal the underlying state. We then do this for the informed agents time-$t$ expectation of the uninformed's $t+1$ and $t+2$ forecast error and calculate the share of informed agents necessary to fully reveal the underlying state. We repeat this process, removing all higher-order belief dynamics sequentially. The share of informed agents that can exist in the model before perfect revelation occurs roughly doubles as all HoBs are removed. This suggests that higher-order beliefs play a crucial role in information transmission.

The final section of the paper shows how our results can be mapped into a well-known asset pricing framework. Traders with constant absolute risk aversion utility observe two types of Gaussian shocks—shocks that are common knowledge and those observed with idiosyncratic noise. The idiosyncratic noise can be interpreted as traders living on Lucas Islands (Lucas (1972)) with incomplete information sharing. We demonstrate how our results can shed light on asset pricing anomalies.

**Contacts with Literature.** To the best of our knowledge, there is no result equivalent to Theorem 1 despite the fact that signal extraction problems have played a foundational role in many literatures. Muth (1960) lays the groundwork for rational expectations and provides a formula for solving a signal extraction problem where a permanent and transitory shock cannot be disentangled. Kydland and Prescott (1982) contains a similar signal extraction problem embedded into a business cycle framework. The inability to disentangle permanent from transitory shocks in Friedman's (1957) life cycle permanent income theory leads to the information aggregation bias of Goodfriend (1992) and Pischke (1995). Lucas (1972), Mills (1982) and Wallace (1992) build monetary frameworks with signal extraction as the linchpin. More recently, connections have been made between the typical signal extraction problem and other informational frictions, like rational inattention (Luo and Young (2014)); and the apparent tradeoff between signal processing and discounting (Gabaix and Laibson (2022)).

Theorem 1 is likely operational in many dynamic models. Several recent papers study similar forms of dispersed information in dynamic macro or asset pricing models. In addition to the papers listed above, a non-exhaustive list includes, Hellwig and Venkateswaran (2009), Lorenzoni (2009), Mackowiak and Wiederholt (2009), Angeletos and La'O (2009), Angeletos and La'O (2013), and Huo and Pedroni (2023). Angeletos and Lian (2016) provides an excellent review of incomplete information in macro modeling. Our theorem therefore presents a class of rational expectations equilibria that could potentially emerge in such models, but have yet to be characterized. The theorem also provides useful decompositions that facilitate interpretation. Albeit speculative, we believe there many economic frameworks in which our theorem could be operational and potentially helpful. We provide a mapping into a well-known asset pricing model in Section 3.3.

Our approach to solving rational expectations models with dispersed information relies on finding fundamental moving average (FMAs) representations, applying the Wiener-Kolmogorov optimal prediction formula, and solving for a rational expectations equilibrium via analytic functions. Hansen and Sargent (1980) and Townsend (1983a) were early advocates of deriving FMAs as a way of finding the agents' innovations representation. Like our paper, Taub (1989), Kasa (2000), Walker (2007), Rondina (2009), Acharya (2013), Kasa et al. (2014), Rondina and Walker (2021), Mao et al. (2021), Huo and Takayama (2022), Jurado (2023), Han et al. (2023) and Chahrour and Jurado (2023) employ frequency domain techniques to solve for a rational expectations equilibrium with some form of information friction. Our solution procedures rely on analytic function theory first introduced by Whiteman (1983).[2] Futia (1981) was perhaps the earliest adopter of such methods in models with heterogeneous agents. While we argue our approach is the most straightforward path to an analytic solution, the methodology is not crucial for the result.

Equivalence results have been employed in the incomplete information literature to great effect. For example, needing to find a way to compact a potentially infinite dimensional state space, Sargent (1991)

[2] Readers unfamiliar with these techniques can consult Appendix B where we solve representative agent versions of the models and discuss the solution procedure in more detail.

first recognized that low-order ARMA processes could mimic infinite-dimensional moving average representations. Kasa (2000) and Chahrour and Jurado (2023) pushed this interpretation further by showing the ease with which these calculations can be done using complex analysis. More recently, Huo and Pedroni (2020) and Angeletos and Huo (2021) are two excellent examples of how equivalence results can aid in computation, interpretation, and evaluation of equilibria with information distortions.

## 2 A HIERARCHICAL-DISPERSED EQUIVALENCE

We first introduce a model with hierarchical information in which there are two agent types—informed and uninformed. Uninformed agents can learn through endogenous variables, which makes the equilibrium non-trivial. We then solve for an economy with a continuum of agents who each observe an idiosyncratic signal about the true underlying state of the economy. We derive our equivalence, in the aggregate, by setting the signal-to-noise ratio in the dispersed economy equal to the share of informed agents in the hierarchical economy. The usefulness of our result is analyzed in the following section, where we study objects (e.g., higher-order beliefs) that are tractable in the hierarchical environment and can be mapped directly into dispersed-informational settings. Before turning to these results, we establish the full and incomplete information homogeneous beliefs equilibria that serve as limiting cases in our

### 2.1 HIERARCHICAL INFORMATION

There are two distinct groups of agents; the first group, in proportion $\mu$, observes the underlying shocks directly, $\Omega^I = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$. This group is fully informed ($I$) and does not solve a signal extraction problem. The second group, in proportion $1-\mu$, only observes the sequence of the endogenous variable, $\Omega_t^U = \{y_{t-j}\}_{j=0}^{\infty}$ and are uninformed ($U$). The corresponding model to be solved is given by

$$y_t = \beta\mu\mathbb{E}^I\left(y_{t+1}|\{\varepsilon_{t-j}\}_{j=0}^{\infty}\right) + \beta(1-\mu)\mathbb{E}^U\left(y_{t+1}|\{y_{t-j}\}_{j=0}^{\infty}\right) + x_t \tag{2}$$

$$x_t = A(L)\varepsilon_t \tag{3}$$

where $x_t = A(L)\varepsilon_t = A_0\varepsilon_t + A_1\varepsilon_{t-1} + \cdots$, $L$ is a lag operator $Lx_t \equiv x_{t-1}$, and the coefficients satisfy square summability, $\sum_j A_j^2 < \infty$. Representation (3) places no restrictions on the serial correlation properties of $x_t$. The Wold Decomposition Theorem allows for such a general representation.

We need to ensure that the endogenous variable, $y_t$, does not fully reveal the underlying shocks to the uninformed agents in equilibrium. One manner to impose that the equilibrium exists in a subspace of $\varepsilon_t$ is to assume the endogenous variable is given by[3]

$$y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t \qquad |\lambda| \in (-1,1) \tag{4}$$

[3]This particular type of signal extraction problem was first encountered in a rational expectations setting in the seminal work of Townsend (1983b) and is motivated further in Rondina and Walker (2021). The appendix of Rondina and Walker (2021) discusses the extension of including multiple parameters inside the unit circle, $y_t = (L-\lambda_1)(L-\lambda_2)\widetilde{Y}(L)\varepsilon_t$, and makes the case that this alternative specification would not alter any results contained herein.

If $|\lambda| \in (-1,1)$, then agents only observing the sequence $\{y_{t-j}\}_{j=0}^{\infty}$ will not be able to infer the underlying shocks, $\{\varepsilon_{t-j}\}_{j=0}^{\infty}$ and the variance of their forecast errors will be larger relative to the agents who observe the underlying shocks, $\{\varepsilon_{t-j}\}_{j=0}^{\infty}$. Using the terminology of Rozanov (1967), the $y_t$ process is not fundamental for the $\varepsilon_t$ sequence, and thus the information set generated by observing the $y_t$'s is a strict subset of that generated by the $\varepsilon_t$'s.

To understand this endogenous signal extraction problem, first consider a similar *exogenous* signal extraction problem

$$s_t = (L-\vartheta)\varepsilon_t = -\vartheta\varepsilon_t + \varepsilon_{t-1}, \tag{5}$$

where $\varepsilon_t$ is a mean-zero, normally distributed variable with variable $\sigma_\varepsilon^2$ and $\vartheta \in (0,1)$. Rondina and Walker (2021) show the mean-squared error minimizing prediction for $\varepsilon_t$ conditional on observing current and past $s$ is

$$\mathbb{E}\left(\varepsilon_t | \{s_{t-j}\}_{j=0}^{\infty}\right) = \underbrace{\vartheta^2\varepsilon_t}_{\text{information}} \quad - \quad \underbrace{(1-\vartheta^2)[\vartheta\varepsilon_{t-1} + \vartheta^2\varepsilon_{t-2} + \vartheta^3\varepsilon_{t-3} + \cdots]}_{\text{noise from confounding dynamics}}. \tag{6}$$

information + noise from confounding dynamics

Expression (6) suggests that the process (5) is informationally equivalent to a noisy signal about $\varepsilon_t$, where the noise is the linear combination of past shocks, and the signal-to-noise ratio is measured by $\vartheta^2$. A $\vartheta$ closer to zero results in less information and more noise but, at the same time, it also makes past shocks less persistent. As $\vartheta \to 0$, there is no information in $s_t$ about $\varepsilon_t$ and the optimal prediction is 0, the unconditional average. As long as $|\vartheta| \in (-1,1)$, the value of $\varepsilon_t$ will *never* be learned and in this sense, the *history* of the fundamental shock acts as a noise shock. The shocks are perfectly correlated and no super-imposed noise process is necessary to keep full revelation of information from occurring. An infinite history of past shocks is not sufficient because the dynamic history of the shock confounds agents into making forecast errors that would be persistent under the standard full-information rational expectations case. Because of this, Rondina and Walker (2021) refer to this type of noise as displaying *confounding dynamics.*

Moreover, signal extraction problems (signal plus noise) can be calibrated to contain the same information as a stochastic process with confounding dynamics. Specifically, suppose that agents observe an infinite history of the signal

$$\mathscr{S}_t = \varepsilon_t + \eta_t, \tag{7}$$

where $\eta_t \overset{iid}{\sim} N\left(0,\sigma_\eta^2\right)$. The optimal expectation is given by $\mathbb{E}(\varepsilon_t|\mathscr{S}^t) = \tau\mathscr{S}_t$, where $\tau$ is the relative weight given to the signal, $\tau = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$. Appendix A proves the following equivalence between the two signal extraction problems, where equivalence is defined as equality of variance of the forecast error conditioned on the infinite history of the observed signal,

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|s^t\right)\right)^2\right] = \mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|\mathscr{S}^t\right)\right)^2\right] \tag{8}$$

when

$$\vartheta^2 = \tau = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\eta^2} \tag{9}$$

Notice that when the signal-to-noise ratio increases (decreases), this corresponds to a higher (lower) absolute value of $\vartheta$. In the limit, as $\sigma_\eta^2 \to 0$, then $\vartheta^2 \to 1$ ensures exact recovery of the state in both cases. Thus, we could have an environment with more shocks than signals to preserve incomplete information in equilibrium or we can impose a non-fundamental process such as (4).

Returning to our endogenous signal extraction problem, we must first find the corresponding innovations associated with observing current and past $y_t$; thus, we must flip the $\lambda$ root from inside the unit circle to outside the unit circle without changing the moments of the $y_t$ process. This transformation is accomplished through the use of Blaschke factors, $\mathscr{B}_\lambda(L) \equiv (L-\lambda)/(1-\lambda L)$

$$y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t = (1-\lambda L)\widetilde{Y}(L)e_t \tag{10}$$

$$e_t = \left(\frac{L-\lambda}{1-\lambda L}\right)\varepsilon_t = (L-\lambda)(\varepsilon_t + \lambda\varepsilon_{t-1} + \lambda^2\varepsilon_{t-1} + \cdots) \tag{11}$$

Note that we are operating in well-defined Hilbert spaces with the covariance generating function serving as the modulus and that Blaschke factors have a modulus of one, $\mathscr{B}_\lambda(z)\mathscr{B}_\lambda(z^{-1}) = 1$, supporting the equality in (10). Note also that *conditional* expectations differ in the $e_t$ and $\varepsilon_t$ spaces.

The guess of the equilibrium process (10) must be verified, and uniquely so. In order to prove that an equilibrium of the form (4) exists, we need to derive a restriction on the exogenous process $(x_t)$. Merely assuming that the exogenous process is not invertible is insufficient. In a heterogeneous agent setup, the informed agents will impound information into the sequence of endogenous variables and uninformed agents will engage in *endogenous* signal extraction. We allow the less informed agents to learn through observations of the endogenous variable, and therefore need to prove that the equilibrium process will not reveal the underlying shocks perfectly. The following proposition derives a necessary restriction to keep the uninformed from learning the fundamental shocks and characterizes the unique rational expectations equilibrium.

**Proposition 1.** *Consider the economy described by* (2)—(4)*. If* $\beta \in (0,1)$ *and there exists a* $|\lambda| \in (-1,1)$ *such that*

$$A(\lambda) - \frac{\mu\beta A(\beta)}{\mu\lambda + (1-\mu)\left(\frac{\beta-\lambda}{1-\lambda\beta}\right)} = 0 \tag{12}$$

*then the unique rational expectations equilibrium is given by*

$$y_t = \frac{1}{L-\beta}\left\{LA(L) - \beta A(\beta)\left(\frac{\mu\lambda + (1-\mu)\mathscr{B}_\lambda(L)}{\mu\lambda + (1-\mu)\mathscr{B}_\lambda(\beta)}\right)\right\}\varepsilon_t \tag{13}$$

*with* $\mathscr{B}_\lambda(L) \equiv \frac{L-\lambda}{1-\lambda L}$ *and* $\mathscr{B}_\lambda(\beta) \equiv \frac{\beta-\lambda}{1-\lambda\beta}$.

*Proof.* See Appendix A. □

The intuition behind Proposition 1 is as follows: The initial guess of $y_t = (L-\lambda)Y(L)\varepsilon_t$ with $|\lambda| < 1$ implies uninformed agents, through knowledge of the endogenous variable alone, will be able to infer the linear combination of current and past $e_t = \mathscr{B}_\lambda(L)\varepsilon_t$. In order for this informational assumption to survive in equilibrium, it must be the case that knowledge of the model does not provide any *additional* information. More precisely, through knowledge of the structural model (2), uniformed agents are able to subtract off their expectation ($\mathbb{E}^U$) from the equilibrium. What remains is the expectation of the informed ($\mathbb{E}^I$) and the exogenous process, $x_t$. That is,

$$
\begin{aligned}
y_t - \beta(1-\mu)\mathbb{E}^U(y_{t+1}) &= \beta\mu\mathbb{E}^I(y_{t+1}) + x_t \\
&= \beta\mu L^{-1}\left[(L-\lambda)Y(L) - \frac{\lambda A(\beta)}{h(\beta)}\right]\varepsilon_t + A(L)\varepsilon_t \qquad (14)
\end{aligned}
$$

where the last equality follows from the proof of Proposition 1 in Appendix A. Equation (14) provides the exact linear combination of structural shocks that the uninformed agents are able to glean from performing endogenous signal extraction. The information provided by (14) must be equivalent to $e_t$ in order for equilibrium to be consistent with rational expectations. This will be true if and only if (14) vanishes at $L = \lambda$. Condition (12) ensures that this is the case.

The equilibrium representation of Proposition 1 is algebraically the cleanest because it makes clear that as the share of informed agents goes to zero, $\mu \to 0$, the economy converges to the homogeneous beliefs, incomplete information equilibrium. Conversely, as the share of informed agents approaches one, $\mu \to 1$, the equilibrium converges to the full information equilibrium. We state this as a corollary.

**Corollary 1.** *Consider the economy described by (2)–(3). For $\mu = 0$, if the exogenous process satisfies the restriction*

$$
A(\lambda) = 0 \qquad (15)
$$

*with $|\lambda| \in (-1,1)$, then the unique rational expectations equilibrium is given by*

$$
\begin{aligned}
y_t &= \left(\frac{L(1-\lambda L)\widetilde{A}(L) - \beta(1-\lambda\beta)\widetilde{A}(\beta)}{L-\beta}\right)e_t \qquad (16)\\
e_t &= \left(\frac{L-\lambda}{1-\lambda L}\right)\varepsilon_t
\end{aligned}
$$

*If $\mu = 1$ or $|\lambda| > (-1,1)$, then the rational expectations equilibrium is unique and given by*

$$
y_t = \left(\frac{LA(L) - \beta A(\beta)}{L-\beta}\right)\varepsilon_t \qquad (17)
$$

The full-information equilibrium (17) is the ubiquitous Hansen-Sargent formula [Hansen and Sargent (1980)]. 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 that we take advantage of throughout the paper. The first component, $LA(L)/(L-\beta)$, is the perfect foresight equilibrium; that is, assuming a representative agent, solve the model (2) forward, recursively substituting, imposing the law of iterated expectations and applying a no-bubble condition to arrive at

$$y_t = \mathbb{E}_t \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t \left( \frac{LA(L)}{L-\beta} \right) \varepsilon_t \tag{18}$$

If we appended the agent's information set with future values of $\varepsilon_t$, such that agents have perfect foresight (PF) $\Omega_t^{PF} = \{\varepsilon_{t-j}\}_{j=-\infty}^{\infty}$, (18) (after removing the expectation operator) would be the rational expectations equilibrium. Therefore the last element of the H-S formula, $\beta A(\beta)/(L-\beta)$, represents the conditioning down associated with only observing current and past $\varepsilon_t$'s. Subtracting off this precise linear combination of future shocks, $\beta A(\beta)\sum_j \beta^j \varepsilon_{t+j}$, stems from knowledge that the model is given by (18) and the information set of $\Omega_t^I = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$.[4]

As opposed to (13), the following corollary shows that there are equivalent representations that have a more natural economic interpretation.

**Corollary 2.** *The equilibrium described in Proposition 1 has an equivalent representation in e space given by*

$$y_t = \frac{1}{L-\beta}\Big\{(1-\lambda L)LH^U(L) - (1-\lambda\beta)\beta H^U(\beta)\Big\}e_t, \tag{19}$$

*where* $H^U(L) = (L-\lambda)^{-1}\{x_t - \mu\beta[y_{t+1} - \mathbb{E}_t^I(y_{t+1})]\}$.
*And a representation in* $\varepsilon$ *space given by*

$$y_t = \frac{1}{L-\beta}\Big\{LH^I(L) - \beta H^I(\beta)\Big\}\varepsilon_t \tag{20}$$

*where* $H^I(L) = x_t - (1-\mu)\beta[y_{t+1} - \mathbb{E}_t^U(y_{t+1})]$

*Proof.* Follows directly from Proposition 1. □

In models with heterogeneous beliefs, optimal expectations imply that agents must take into consideration the actions of others. The following representation of equilibrium shows how the agents of this model extract information from other agents' forecast *errors* in forming their beliefs of market fundamentals. Representations (19) and (20) demonstrates how agents' beliefs about market fundamentals are intricately tied to the beliefs of other agents. For the informed (uninformed) agents, the market fundamental is a combination of the exogenous process, $x_t$, *and* the forecast error of the uninformed (informed) agents. Due to these speculative dynamics, agents take into account the forecast error of the other agent type when formulating their belief for market fundamentals. Using these corollaries, we derive an analytical form of these higher-order beliefs in Section 3.1. Before doing so, we derive an equivalence between this economy and a dispersed information setup.

### 2.2 DISPERSED INFORMATION

In this section, we study equilibria in which each agent observes its own particular "window of the world," as in Phelps (1969). Agents observe a noisy signal of the innova-

[4]As shown in Appendix A of Hansen and Sargent (1980), agents who know the model is given by (18) will form expectations optimally by subtracting off the principal part of the Laurent series expansion of $A(L)$ around $\beta$, which is $\beta A(\beta)/(L-\beta)$.

tion which is idiosyncratic across agents. Information is dispersed in the sense that, although complete knowledge of the fundamentals is not given to any one agent, by pooling the noisy signal of all agents, it is possible to recover the full information equilibrium.

Specifically, we assume agents (indexed by $i$) observe the sequence of current and past endogenous variables $\{y_{t-j}\}_{j=0}^{\infty}$ in addition to a sequence of noisy signals, specified as $\varepsilon_{it} = \varepsilon_t + v_{it}$ with $v_{it} \sim N\left(0, \sigma_v^2\right)$ for $i \in [0,1]$ and $\Omega_t^i = \{y_{t-j}, \varepsilon_{i,t-j}\}_{j=0}^{\infty}$ for $i \in [0,1]$. The model to be solved is

$$y_t = \beta \int_0^1 \mathbb{E}^i [y_{t+1} | \Omega_t^i] di + x_t \tag{21}$$
$$x_t = A(L)\varepsilon_t$$

When the noise is driven to zero, $\sigma_v^2 \to 0$, this setup is equivalent to a full information equilibrium.

What is unique about this setup is that each agent formulates a forecast by extracting optimally the information from a vector of two signals $\left(y_t, \varepsilon_{it}\right)$. The idea of deriving a fundamental representation developed in the hierarchical case extends naturally to a multivariate setting. The mapping between the signal and innovations is now a matrix, and the invertibility of that matrix determines the information content of the signals. The mapping is given by

$$\left(\begin{array}{c} \varepsilon_{it} \\ y_t \end{array}\right) = \left[\begin{array}{cc} 1 & 1 \\ (L-\lambda)\, Y(L) & 0 \end{array}\right] \left(\begin{array}{c} \varepsilon_t \\ v_{it} \end{array}\right) \tag{22}$$

Given the candidate price function, this matrix is of rank 1 at $L = \lambda$ and so it cannot be inverted. As shown in Appendix A and Rondina (2009), the invertible representation is derived through use of a combination of Blaschke factors and factorization of the signal $\varepsilon_{it}$. The optimal expectation will always be given by the sum of two terms: a linear combination of current and past innovations $\varepsilon_t$ and a linear combination of current and past idiosyncratic noise $v_{it}$. Appendix A shows that taking the average of the expectations across agents, the second term will be zero, yielding

$$\overline{\mathbb{E}}_t(y_{t+1}) = [(L-\lambda)Y(L) + \lambda Y_0]\left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2}\right)\varepsilon_t + [(1-\lambda L)Y(L) - Y_0]\left(\frac{\sigma_v^2}{\sigma_\varepsilon^2+\sigma_v^2}\right)e_t \tag{23}$$

Substituting this expectation into the equilibrium and solving gives the following proposition.

**Proposition 2.** *Consider the economy described by* (21). *If* $\beta \in (0,1)$ *and there exists a* $|\lambda| \in (-1,1)$ *such that*

$$A(\lambda) - \frac{\tau \beta A(\beta)}{\tau\lambda + (1-\tau)\left(\frac{\beta-\lambda}{1-\lambda\beta}\right)} = 0 \tag{24}$$

*then the unique rational expectations equilibrium is given by*

$$y_t = \frac{1}{L-\beta}\left\{LA(L) - \beta A(\beta)\left(\frac{\tau\lambda + (1-\tau)\mathscr{B}_\lambda(L)}{\tau\lambda + (1-\tau)\mathscr{B}_\lambda(\beta)}\right)\right\}\varepsilon_t \tag{25}$$

*with* $\mathscr{B}_\lambda(L) \equiv \frac{L-\lambda}{1-\lambda L}$, $\mathscr{B}_\lambda(\beta) \equiv \frac{\beta-\lambda}{1-\lambda\beta}$ *and* $\tau = \sigma_\varepsilon^2/(\sigma_v^2 + \sigma_\varepsilon^2)$ *is the signal-to-noise ratio.*

*Proof.* See Appendix A. □

The main result of the section is that the rational expectations equilibrium under dispersed information takes the same form as the equilibrium under hierarchical information (13), once the parameter that governs the share of informed agents $\mu$ is appropriately reinterpreted. This analogous representation allows one to immediately apply the characterizations of the previous section—and the implications discussed in the next section—to the more realistic dispersed information setup. At the same time, since no agent is alike in the dispersed information setup, there are aspects of the equilibrium that will not emerge in the hierarchical case.
Theorem 1 follows immediately.

**Theorem 1.** *Let* $\tau \equiv \sigma_\varepsilon^2/(\sigma_v^2 + \sigma_\varepsilon^2)$ *be the signal-to-noise ratio in the dispersed-information economy and* $\mu$ *be the share of informed agents in the hierarchical-information economy. The rational expectations equilibrium of Proposition 2 is equivalent to the rational expectations equilibrium of Proposition 1 when* $\mu = \tau$.

The theorem states that in terms of aggregates, the dispersed information setup is identical (i.e., same existence condition (12) is equal to (24), and same equilibrium function (13) is equal to (25)) to the hierarchical information equilibrium when the signal-to-noise ratio $\tau \equiv \sigma_\varepsilon^2/(\sigma_v^2 + \sigma_\varepsilon^2)$ is equal to the proportion of informed traders, $\mu$. This equivalence result can be understood by thinking of the strategic behavior of the dispersedly informed agent. Each agent $i$ receives a privately observed signal $\varepsilon_{it}$ and a publicly observed signal $y_t$ about the unobserved fundamental $\varepsilon_t$. The optimal behavior—in terms of forecast error minimization—is for the agent to act *as if* the signal $\varepsilon_{it}$ contains no noise and thus is equal to the true state $\varepsilon_t$, in measure proportional to the informativeness of the signal $\tau$. At the same time, to act as if the signal is pure noise and thus it would be optimal to ignore it and act just upon the public signal $y_t$, this in measure $(1-\tau) \equiv \sigma_v^2/(\sigma_v^2 + \sigma_\varepsilon^2)$. Thus, in a dispersed information setting each agent behaves optimally by employing a "mixed strategy" approach: act as if they possess the full information of the informed agents $\Omega^I$ with probability $\tau$, and act as if they possess just the public information of the uninformed agents $\Omega^U$ with probability $1-\tau$.

While Theorem 1 guarantees equivalence with the hierarchical setup at the aggregate level, important differences between the two equilibria at the individual agent level remain. First, the dispersed information equilibrium displays a well defined cross-sectional distribution of beliefs, as opposed to the degenerate distribution that would emerge in the hierarchical case. Second, the cross-sectional variation is perpetual in the sense that the unconditional cross-sectional variance is positive. In other words, agents' beliefs are in perpetual disagreement. These two results are stated in the following proposition.

**Proposition 3.** *The cross-section of beliefs of agent* $i$ *are given by*

$$\mathbb{E}_t^i(y_{t+j}) = \mathbb{E}_t^I\left(y_{t+j}\right) - (1-\tau)Y_{j-1}\frac{1-\lambda^2}{1-\lambda L}\varepsilon_t - \tau Y_{j-1}\frac{1-\lambda^2}{1-\lambda L}v_{it} \text{ for } j = 1,2,\dots. \tag{26}$$

*The unconditional variance of the difference in beliefs across agents is given by*

$$\tau^2\left(1-\lambda^2\right)Y_{j-1}^2\sigma_v^2 \text{ for } j=1,2,... \tag{27}$$

*Proof.* See Appendix A. □

Under full information, the beliefs would coincide with the expectation $\mathbb{E}_t^I\left(y_{t+j}\right)$. The difference of the beliefs of agent $i$ with respect to the full information has two components—one is common across agents, one is specific to each agent. The common component is analogous to the error associated with being uninformed and was studied in the previous section, $(1-\tau)Y_{j-1}((1-\lambda^2)/(1-\lambda L))\varepsilon_t$. The second component is the result of the agent acting as informed but not being able to cleanly distinguish between $\varepsilon_t$ and $v_{it}$. Optimal signal extraction implies that this particular linear combination of idiosyncratic shocks will infiltrate agent $i$'s optimal time-$t$ expectation, while aggregating over all agents eliminates this term. Thus, the unconditional variance of beliefs will be positive for all $j$. Proposition 3 offers an analytical form that can be useful in calibrating key parameters of cross-sectional beliefs.

# 3 IMPLICATIONS OF THEOREM 1

The significance of Theorem 1 is that it can be operationalized to facilitate interpretation, which stems from the relative ease with which one can analyze the hierarchical equilibrium and the challenges associated when information is dispersed. Calculating objects like higher-order belief dynamics when information can be ordered is feasible, but much more challenging when information is equally distributed. Our interpretation of Theorem 1—that dispersed traders act of as informed with probability equal to the signal-to-noise ratio and uninformed with the complementary probability—facilitates our analysis.

## 3.1 HIGHER-ORDER BELIEFS

Higher-order beliefs in the hierarchical equilibrium follow most naturally from the equilibrium representations of Corollary 2, which we repeat here for convenience,

$$y_t = \frac{1}{L-\beta}\left\{(1-\lambda L)LH^U(L)-(1-\lambda\beta)\beta H^U(\beta)\right\}e_t,$$

where $H^U(L)=(L-\lambda)^{-1}\{x_t-\mu\beta[y_{t+1}-\mathbb{E}_t^I(y_{t+1})]\}$. And a representation in $\varepsilon$ space given by

$$y_t = \frac{1}{L-\beta}\left\{LH^I(L)-\beta H^I(\beta)\right\}\varepsilon_t$$

where $H^I(L)=x_t-(1-\mu)\beta[y_{t+1}-\mathbb{E}_t^U(y_{t+1})]$. These Hansen-Sargent equations make clear that each agent type believes that market fundamentals (i.e., the stochastic process to be forecast) consists of the underlying exogenous process, $x_t$, and the *forecast error* of the other agent. Agents are forecasting the forecast errors of the other agent type. The restriction from Proposition 1, $A(\lambda)-(\mu\beta A(\beta))/(\mu\lambda+(1-\mu)\mathscr{B}_\lambda(\beta))=0$ ensures that uninformed agents cannot learn more from the informed forecast error than the space spanned by the $e_t$ process.

In order to derive *higher*-order beliefs, we iterate the equilibrium equation forward by one period,

$y_{t+1} = \beta\mu\mathbb{E}^I_{t+1}[y_{t+2}] + \beta(1-\mu)\mathbb{E}^U_{t+1}[y_{t+2}] + x_{t+1}$, noting that the functional form of the equilibrium is $y_t = (L-\lambda)Y(L)\varepsilon_t$; the appendix shows the time $t+1$ average expectation of the endogenous variable at $t+2$ can be written as the actual value at $t+2$ minus the average market forecast error, namely

$$\mu\mathbb{E}^I_{t+1}y_{t+2} + (1-\mu)\mathbb{E}^U_{t+1}y_{t+2} = y_{t+2} + \mu Y_0\lambda\varepsilon_{t+2} - (1-\mu)Y_0\mathscr{B}_\lambda(L)\varepsilon_{t+2} \tag{28}$$

The average market forecast error on the RHS of (28) has two components: the first term represents the error made by the informed agents, $Y_0\lambda\varepsilon_{t+2}$, appropriately weighted by the mass of informed agents in the market, $\mu$; the second term, $Y_0\mathscr{B}_\lambda(L)\varepsilon_{t+2} = Y_0 e_{t+2}$, represents the forecast error of the uninformed agents, weighted by $1-\mu$. We know from the form of the lag polynomial $\mathscr{B}_\lambda(L)$ that the forecast error of uninformed agents contains a linear combination of current and past innovations of the informed agents' information set, $e_{t+2} = (L-\lambda)(1-\lambda L)\varepsilon_{t+2} = (L-\lambda)(\varepsilon_{t+2} + \lambda\varepsilon_{t+1} + \lambda^2\varepsilon_t + \cdots)$. Therefore, the informed agents' time-$t$ expectation of the time $t+1$ average expectation is

$$\mathbb{E}^I_t\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \mathbb{E}^I_t y_{t+2} - (1-\mu)Y_0\left(\frac{1-\lambda^2}{1-\lambda L}\right)\lambda\varepsilon_t \tag{29}$$

Hence, the informed agents will always do better (a smaller forecast error), if they correct their expectation of the average price according to the forecast errors of the uninformed. Conversely, the uninformed form HoBs but the law of iterated expectations holds in their case because the forecast errors of the informed, $Y_0\mu\lambda\varepsilon_{t+2}$, are not forecastable conditional on the uninformed's information set at time $t$, and so $\mathbb{E}^U_t(\overline{\mathbb{E}}_{t+1}y_{t+2}) = \mathbb{E}^U_t y_{t+2}$.

An immediate consequence of informed agents forming HoBs is that the law of iterated expectations fails to hold with respect to the average expectations operator,

$$\overline{\mathbb{E}}_t\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \overline{\mathbb{E}}_t y_{t+2} - \mu(1-\mu)Y_0\left(\frac{1-\lambda^2}{1-\lambda L}\right)\lambda\varepsilon_t \tag{30}$$

As shown in Appendix A, the structure of HoBs at any order can be analytically characterized as

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\cdots\overline{\mathbb{E}}_{t+j}y_{t+j+1} = \overline{\mathbb{E}}_t y_{t+j+1} - (1-\mu)\left(\sum_{i=1}^{j}(\mu\lambda)^i Y_{j-i}\right)\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t$$

This equation shows that higher-order beliefs are a discounted function of structural shocks with discount factor equal to the share of informed agents, $\mu$, and the degree of asymmetric information, as indexed by $\lambda$. We use this equation below to assess the extent to which information is impounded into the equilibrium price via higher-order belief dynamics.

It is optimal for informed agents to adjust expectations by correcting the forecast errors of the uninformed. However, no such informational advantage exists in the dispersed equilibrium. Do agents even form higher-order beliefs? Can they be characterized in closed form? To address these questions, using Theorem 1, we can write the time-$t$ expectation of agent $i$ of the equilibrium at $t+1$ as

$$\mathbb{E}_{it}\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \mu\mathbb{E}_{it}\left(\mathbb{E}^I_{t+1}y_{t+2}\right) + (1-\mu)\mathbb{E}_{it}\left(\mathbb{E}^U_{t+1}y_{t+2}\right)$$

From the hierarchical equilibrium, we know that $\mathbb{E}^U_{t+1} y_{t+2} = \mathbb{E}^I_{t+1} y_{t+2} - Y_0 \frac{1-\lambda^2}{1-\lambda L} \varepsilon_{t+1}$. We also notice that, because the information set of an arbitrary agent $i$ is strictly smaller than the information set of an informed agent of the hierarchical equilibrium and because the law of iterated expectations holds at the single agent level, we have $\mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}^I_{t+1} y_{t+2} = \mathbb{E}_{it} y_{t+2}$. The law of iterated expectations holding at the single agent level also implies $\mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}^U_{t+1} y_{t+2} = \mathbb{E}_{it}\mathbb{E}^U_{t+1} y_{t+2}$. Therefore

$$\mathbb{E}_{it}\left(\overline{\mathbb{E}}_{t+1} y_{t+2}\right) = \mu \mathbb{E}_{it} y_{t+2} + (1-\mu)\mathbb{E}_{it} y_{t+2} - (1-\mu) Y_0 \mathbb{E}_{it}\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_{t+1} \tag{31}$$

Forming higher-order beliefs and breaking the law of iterated expectations follows if the last term is nonzero. Appendix A shows

$$(1-\mu) Y_0 \mathbb{E}_{it}\left(\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}\right) = Y_0 (1-\mu)\left(\frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}\right)\left(\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + \sigma^2_v}\right)(\varepsilon_t + v_{it})$$

Aggregating over all agents provides an equivalence to the hierarchical average expectation. At face value, this result seems counterintuitive because all agents are similarly uninformed. Each agent must think that her information is somehow superior to the information of the other agents in order for the law of iterated expectations to not be applicable. The intuition behind Theorem 1 provides the answer. Take any arbitrary agent $i$. This agent is instructed by the optimality of signal extraction to act as informed with probability $\mu$. In so doing, she will recognize that a fraction $1-\mu$ of agents is contemporaneously acting as uninformed. It follows that as an informed agent, agent $i$ should forecast the forecast error of the agents acting as uninformed and embed it into her expectations about the future. She will adjust her time-$t$ forecast according to the collective ignorance of the uninformed agents (i.e., agents inferring the signal as pure noise). This ignorance accumulates at time $t+1$, $t+2$, etc. and therefore, (31) generalizes to higher orders. At the same time, she is acting as uninformed as well and is part of the portion of $1-\mu$ agents of whom she is forecasting the forecast errors. However, the relevance of her individual forecast error is infinitesimal in this regard and thus irrelevant for her reasoning as informed.

**3.2 INFORMATION TRANSMISSION** 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 (or, due to Theorem 1, the share of informed agents), needed to completely reveal the underlying state. That is, we can use the existence criteria of Propositions 1 and 2, specifically Equation (24), to determine the required $\tau$ or $\mu$ necessary to completely reveal the underlying shock sequence, $\varepsilon^t$. 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.

We begin with Figure 1, which characterizes the dispersed information equilibrium of Proposition 2 in the $(\beta,\theta)$ space for the exogenous process, $x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. (Of course given Theorem 1, this figure also characterizes equilibrium for the hierarchical formulation of Proposition 1.) The figure is built

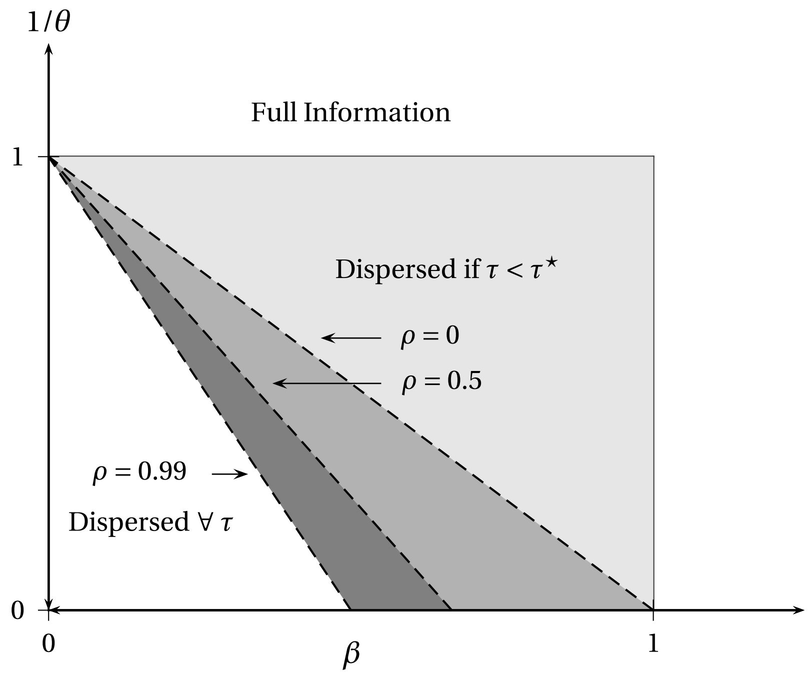


Figure 1: $(\beta, \theta)$ Existence Space. Existence of Dispersed and Full-Information Equilibria following Proposition 2 for $x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Equilibria to the right of the dashed line preserve heterogeneity in information if and only if $\tau < \tau^\star$.

around the following corollary to Proposition 2.

**Corollary 3.** *Consider the dispersed-information economy described by* (21) *of Proposition 2 with* $x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\beta, \rho \in (0,1)$ *and* $\theta > 0$. *The equilibrium is characterized in the* $(\beta, \theta)$ *space according to the following restrictions:*

1. *If* $\theta \le 1$, *a dispersed information equilibrium does not exist and the model is characterized by the full-information equilibrium.*
2. *If* $\theta > 1$, *a dispersed information equilibrium exists for any* $\tau > 0$ *and* $\rho \ge 0$ *if*
$$\theta \ge \left(\frac{1}{1-\beta(1+\rho)}\right) \tag{32}$$
3. *If* $\theta > 1$ *and (32) is not satisfied, a dispersed information equilibrium exists for signal-to-noise ratio* $\tau$ *if and only if* $\tau \in (0, \tau^\star)$ *with*
$$\tau^\star = \frac{(\theta-1)(1-\rho\beta)}{\beta(1+\rho)(1+\theta\beta)}$$

*Proof.* See Appendix A. □

Three points are noteworthy. First, as is evident from Figure 1 and statement 1 of Corollary 3, if $\theta \le 1$, the endogenous variable fully reveals the underlying shock, $\varepsilon_t$, and the equilibrium is consistent with the full information equilibrium. With $\theta \leq 1$, confounding dynamics are not present in the exogenous process and $x_t$ is fundamental for $\varepsilon_t$. 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 signal. The derivations of Section 2.1 demonstrate that an increase in $\theta$ 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 $\beta$ alters the space of existence is similar to that of the serial correlation parameter $\rho$, which is the final point to be made. As the serial correlation in the $x_t$ process increases and $\beta$ increases, it is more difficult to preserve dispersed information, *ceteris paribus* (the dashed line shifts to the left as $\rho$ increases from 0 to 0.99). An increase in $\beta$ and $\rho$ 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.

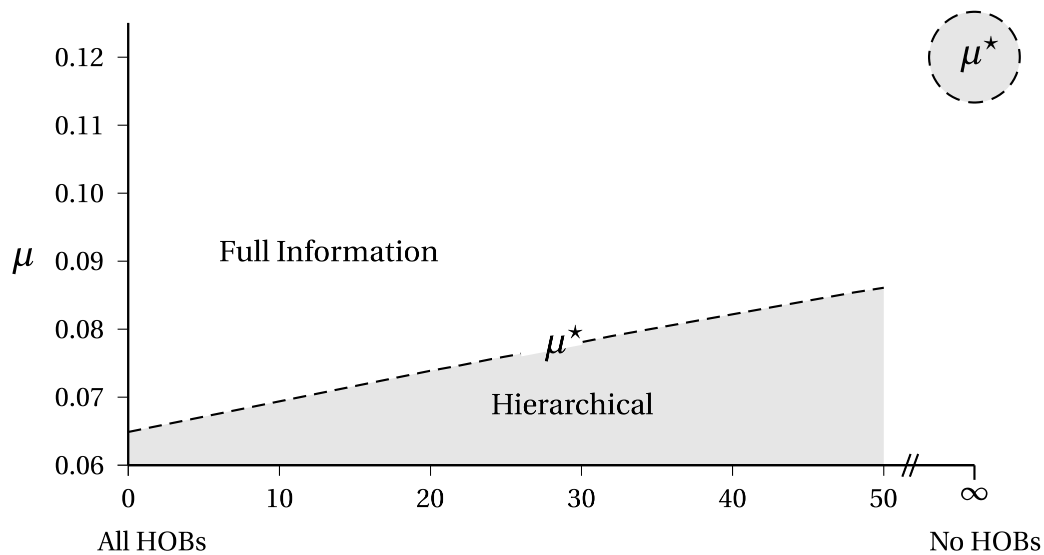


Figure 2: Existence space for the hierarchical information equilibrium as higher-order beliefs are removed from the expectation of informed agents: $x_t = 0.8x_{t-1} + \varepsilon_t + \sqrt{11}\varepsilon_{t-1}$, $\beta = 0.985$.

Higher-order belief dynamics play a crucial role in disseminating information. As discussed above, informed agents are correcting for the bias in the uniformed agents' forecast errors, so there is an important feedback mechanism at work. The uninformed agents are able to extract information about their own forecast errors by observing the endogenous variables due to the formation of HoBs. 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 Propositions 1 and 2 as higher-order belief dynamics are sequentially removed from the expectation of the informed agents. 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 $\infty$ 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.

### 3.3 AN ASSET PRICING EXAMPLE

The purpose of this section is to show how our setup and results can be implemented in a well-known framework. Suppose there is a riskless asset (e.g., Treasury) that pays a constant rate of return $(1+r)$ in perfect elastic supply, and a risky asset (e.g., stock) with price $p_t$ and fundamentals $(f_t)$ that are comprised of a linear combination of two stochastic components. The first component($f_t^I$) is observed without noise. For example, the size of next period's dividend is typically common knowledge among market participants. The second component of fundamentals $(f_t^U)$ is observed with idiosyncratic noise. Our assumption here is that traders are informed about news concerning a particular firm through various media outlets and these outlets do not communicate the news identically to all traders. One interpretation of this assumption is that traders live on Lucas islands [Lucas (1972)] with imperfect information sharing. Traders (indexed by $i$) observe a sequence of noisy signals about unobserved fundamentals $\varepsilon_{it} = \varepsilon_t^U + v_{it}$ with $v_{it} \sim N\left(0, \sigma_v^2\right)$. The information set of the traders is therefore, $\Omega_t^i = \{p_{t-j}, f_{t-j}^I, \varepsilon_{i,t-j}\}_{j=0}^{\infty}$ for $i \in [0,1]$. Fundamentals are given by the convex combination, $f_t = \phi f_t^I + (1-\phi) f_t^U = \phi F^I(L)\varepsilon_t^I + (1-\phi)F^U(L)\varepsilon_t^U$ where $\phi \in (0,1)$ dictates the type of fundamental, and the shocks are assumed Gaussian and orthogonal.

Following the standard CARA-Gaussian calculus,[5] traders submit a linear demand schedule of the form $(1+r) f_t + \mathbb{E}_t^i p_{t+1} - (1+r) p_t$. On the supply side, we follow the standard assumption in the literature by introducing liquidity traders that make the number of shares available to the market a random variable, $u_t$. We assume the stochastic process, $u_t$, is common knowledge and follows an i.i.d. Gaussian

[5]At time $t$, the budget constraint of investor $i$ is given by $w_{i,t+1} = z_{i,t}(p_{t+1} + (1+r) f_t) + (w_{i,t} - z_{i,t} p_t)(1+r)$ where $w_{i,t}$ denotes the wealth of agent $i$ at $t$ and $z_{i,t}$ is the number of units of the risky asset held by agent $i$ at $t$. The investor will choose $z_{it}$ so as to maximize a constant absolute risk aversion utility function $-\mathbb{E}_t^i \exp(-\gamma w_{i,t+1})$, where $\gamma$ is the risk aversion parameter, and $\mathbb{E}_t^i$ denotes the time $t$ conditional expectation of agent $i$. The linear demand function follows from the first-order optimality condition and by assuming stationarity in the conditional variance term. We normalize the risk aversion parameter to unity. While this limits the role of risk aversion, it preserves linearity which allows us to focus on the linear discounted present value model, and permits closed-form solutions.

process. The equilibrium is then

$$p_t = \beta \int_0^1 \mathbb{E}_t^i [p_{t+1} | \Omega_t^i] di + f_t - \beta u_t$$

where $\beta = (1+r)^{-1}$. Defining the average expectation operator as $\bar{\mathbb{E}}_t^0 x_{t+1} = \int_0^1 \mathbb{E}[x_{t+1}|\Omega_t^i] di$, with higher-order values defined as $\bar{\mathbb{E}}_t^k f_{t+k+1} = \bar{\mathbb{E}}_t \bar{\mathbb{E}}_{t+1} \cdots \bar{\mathbb{E}}_{t+k} f_{t+k+1}$, we can write the equilibrium as

$$p_t = \mathbb{E}_t \sum_{k=0}^{\infty} \beta^k (\phi f_{t+k}^I - \beta u_{t+k}) + (1-\phi) \sum_{k=0}^{\infty} \bar{\mathbb{E}}_t^k \beta^k f_{t+k}^U \tag{33}$$

The first summation contains an expectation operator that obeys the law of iterated expectation due to common information across all traders. Setting $\phi = 1$ and defining $f_t - \beta u_t \equiv j_t = J(L)\xi_t$ gives the homogeneous-information equilibrium

$$p_t = \left( \frac{LJ(L)}{L-\beta} - \frac{\beta J(\beta)}{L-\beta} \right) \xi_t = p_t^{\star} - p_t^R \tag{34}$$

For $\phi \in (0,1)$, the heterogeneous information component of fundamentals, $(1-\phi) f_t^U$, becomes operational. Note that this information structure is consistent with our dispersed information setup of Section 2 and therefore the solution is given by Proposition 2. Moreover, Theorem 1 states that we can employ *any* equilibrium representation established under hierarchical information derived in Section 2.1, under the restriction that the signal-to-noise ratio is set equal to the share of informed agents ($\mu = \tau$).

Consider the hierarchical equilibrium representation given by

$$p_t^U = \left( \frac{LF^U(L)}{L-\beta} \right) \varepsilon_t^U - \left( \frac{\beta F^U(\beta)}{L-\beta} \right) \varepsilon_t^U - \left( \frac{\kappa}{1-\lambda L} \right) \varepsilon_t^U \tag{35}$$

where $\kappa = ((1-\mu)(1-\lambda^2)\beta F^U(\beta))/(\beta - \lambda - \mu\beta(1-\lambda^2))$ isolates the term that distinguishes full information from heterogeneous information. The last term on the right-hand side of (35) is useful for showing how variance bound violations can seemingly occur in heterogeneous-information economies (see e.g., Miao et al. (2021), Kasa et al. (2014)). The intuition comes from writing the price as $p_t = p_t^{\star} + p_t^R$, where $p_t^{\star}$ is the perfect foresight price and $p_t^R$ is the conditioning down term due to agents not observing future values of shocks. As first documented in Shiller (1981), a variance inequality can be established from (34) by noting $\text{var}(p_t^{\star}) = \text{var}(p_t) + \text{var}(p_t^R) + 2\text{cov}(p_t, p_t^R) = \text{var}(p_t) + \text{var}(p_t^R) > \text{var}(p_t)$. With homogeneous information, the covariance between $p_t^R$ and $p_t$ is zero. We can write $p_t = \mathbb{E}_t p_t^{\star}$, and $p_t^R$ represents the optimal conditioning down from the perfect foresight equilibrium; hence, $p_t$ and $p_t^R$ will not be correlated due to the orthogonality of optimal prediction and the inequality is established. To see this explicitly, we use the fact that $x^n - y^n = (x-y)(x^{n-1} + x^{n-2}y + x^{n-3}y^2 + \cdots + x^2 y^{n-3} + x y^{n-2} + y^{n-1})$ to write the equilibrium (34) as

$$
\begin{aligned}
p_t &= \left( \frac{L(J_0 + J_1 L + J_2 L^2 + \cdots) - \beta(J_0 + J_1 \beta + J_2 \beta^2 + \cdots)}{L - \beta} \right) \xi_t \\
&= (J_0 + J_1(L+\beta) + J_2(L^2 + L\beta + \beta^2) + J_3(L^3 + L^2\beta + L\beta^2 + \beta^3) + \cdots)\xi_t \\
&= \left( J(\beta) + L(J(\beta) - J_0)\beta^{-1} + L^2(J(\beta) - J_0 - J_1)\beta^{-2} + \cdots \right) \xi_t
\end{aligned}
$$

This representation makes clear that the equilibrium is only a function of current and past shocks, while the term $p_t^R$ is only a function of future shocks, $p_t^R = \left(\frac{\beta J(\beta)}{L-\beta}\right)\xi_t = \beta j(\beta)\sum_{j=0}^{\infty}\beta^j \xi_{t+j+1}$ and therefore $\text{cov}(p_t^R, p_t) = 0$.

Contingent on the econometrician's information set,[6] the heterogeneous-information equilibrium need not obey the zero-correlation property, i.e., $\text{cov}(p_t^R, p_t) \neq 0$. An econometrician who ignores heterogeneity will not properly account for the last term on the right-hand side of (35), which is clearly a function of *past* shocks $\kappa \sum_j \lambda^j \varepsilon_{t-j}^U$. The econometrician will find a non-negative covariance term between the homogeneous-information equilibrium price, $p_t^I$ and the remainder term, $p_t^R$ given by

$$
\text{cov}(p_t^I, p_t^R) = \sigma_{\varepsilon^U}^2 \kappa \left( F^U(\beta) + (\lambda/\beta)(F^U(\beta) - F_0) + (\lambda/\beta)^2(F^U(\beta) - F_0 - F_1) + \cdots \right) \neq 0. \tag{36}
$$

As the share of informed ($\mu$) or the signal-to-noise ratio ($\tau$) goes to one or as the equilibrium becomes fully revealing ($\lambda = 1$), $\kappa$ goes to zero, as does this correlation. However, if there is a meaningful share of uninformed agents, this non-zero correlation can be substantial and used to overturn representative agent results on excess volatility and asset price momentum.

## 4 CONCLUDING COMMENTS

While our results are derived in a univariate framework for transparency, the solution procedures in Rondina and Walker (2021) are a guide to multivariate extensions. The real business cycle model contained therein pushes the limits of our closed-form expressions but also demonstrates that our propositions and corollaries are applicable in much larger models. This paper is not one of limiting cases.

Perhaps more importantly, Theorem 1 can be applied broadly to many models, even when analytical tractability is no longer feasible. As long as the information structure consists of a continuum of agents that receive idiosyncratic signals on the true underlying state, the intuition of Theorem 1 can be invoked. Agents will apply the optimal mixed strategy to signal extraction that can be mapped directly into an informed-uninformed framework.

[6] The qualifying part of the previous sentence is non-trivial; Kasa et al. (2014) derive the Wold representation of the econometrician for a specific information structure and show how violations of the variance bound are possible.

## REFERENCES

**Acharya, Sushant**, "Dispersed Beliefs and Aggregate Demand Management," February 2013. University of Maryland Working Paper.

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

**__ and Jennifer La'O**, "Noisy Business Cycles," *NBER Macroeconomics Annual*, 2009, *24*, 319–378.

**__ and __** , "Sentiments," *Econometrica*, 2013, *81* (2), 739–779.

**__ and Zhen Huo**, "Myopia and Anchoring," *American Economic Review*, April 2021, *4*, 116–1200.

**Chahrour, Ryan and Kyle Jurado**, "Revisiting the Forecasts of Others," October 2023.

**Friedman, Milton**, "The Permanent Income Hypothesis," in "A Theory of the Consumption Function," Princeton University Press, 1957, pp. 20–37.

**Futia, Carl A.**, "Rational Expectations in Stationary Linear Models," *Econometrica*, 1981, *49* (1), 171–192.

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

**Goodfriend, Marvin**, "Information-Aggregation Bias," *American Economic Review*, 1992, *82* (3), 508–519.

**Han, Zhao, Xiaohan Ma, and Ruoyun Mao**, "The Role of Dispersed Information in Inflation and Inflation Expectations," *Review of Economic Dynamics*, 2023, *48*, 72–106.

**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, Christian and Venky Venkateswaran**, "Setting the Right Prices for the Wrong Reasons," *Journal of Monetary Economics*, 2009, *56*, S57–S77.

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

**__ and __** , "Dynamic Information Aggregation: Learning from the Past," *Journal of Monetary Economics*, 2023.

**__ and Naoki Takayama**, "Rational Expectation Models with Higher-Order Beliefs," 2022. Working Paper, Yale University.

**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* (3), 1137–1163.

**Kydland, Finn and Edward C. Prescott**, "Time to Build and Aggregate Fluctuations," *Econometrica*, 1982, *50*, 1345–1370.

**Lorenzoni, Guido**, "A Theory of Demand Shocks," *American Economic Review*, 2009, *99* (5), 2050–84.

**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.

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

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

**Mills, Terence C.**, "Signal Extraction and Two Illustrations of the Quantity Theory," *American Economic Review*, 1982, *72* (5), 1162–1168.

**Muth, John F.**, "Optimal Properties of Exponentially Weighted Forecasts," *Journal of American Statistical Association*, 1960, *55*, 299–306.

**Phelps, Edmund S.**, "The New Microeconomics in Inflation and Employment Theory," *American Economic Review*, 1969, *59* (2), 147–160.

**Pischke, Jorn-Steffen**, "Individual Income, Incomplete Information, and Aggregate Consumption," *Econometrica*, 1995, *63* (4), 805–840.

**Rondina, Giacomo**, "Incomplete Information and Informative Pricing," 2009. Working Paper. UCSD.

— **and Todd B. Walker**, "Confounding Dynamics," *Journal of Economic Theory*, 2021, *196* (September).

**Rozanov, Yu. A.**, *Stationary Random Processes*, San Francisco: Holden-Day, 1967.

**Sargent, Thomas J.**, "Equilibrium with Signal Extraction from Endogenous Variables," *Journal of Economic Dynamics and Control*, 1991, *15*, 245–273.

**Shiller, Robert J.**, "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 1981, *71* (3), 421–436.

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

**Townsend, Robert M.**, "Equilibrium Theory with Learning and Disparate Expectations: Some Issues and Methods," in Edmund S. Phelps and Roman Frydman, eds., *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*, Cambridge University Press, 1983, pp. 169–202.

__ , "Forecasting the Forecasts of Others," *Journal of Political Economy*, 1983, *91* (4), 546–588.

**Walker, Todd B.**, "How Equilibrium Prices Reveal Information in Time Series Models with Disparately Informed, Competitive Traders," *Journal of Economic Theory*, 2007, *137* (1), 512–537.

**Wallace, Neil**, "Lucas's Signal-Extraction Model: A Finite State Exposition with Aggregate Real Shocks," *Journal of Monetary Economics*, 1992, *30* (December), 433–447.

**Whiteman, Charles**, *Linear Rational Expectations Models: A User's Guide*, Minneapolis: University of Minnesota Press, 1983.

**Whittle, Peter**, *Prediction and Regulation by Linear Least-Square Methods*, Minneapolis: University of Minnesota Press, 1983.

# A PROOFS

**A.1 PROOF OF EQUATIONS (8)–(9)** We need to show that the representations (5) and (7) are equivalent in terms of unconditional forecast error variance

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|\mathscr{S}^t\right)\right)^2\right] = \mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|s^t\right)\right)^2\right] \tag{37}$$

when $\vartheta^2 = \tau = \sigma_\varepsilon^2/(\sigma_\varepsilon^2+\sigma_\eta^2)$.

The optimal forecast $\mathbb{E}[\varepsilon_t|\mathscr{S}^t]$ is given by weighting $\mathscr{S}_t$ according to the relative variance of $\varepsilon$, $\mathbb{E}(\varepsilon_t|\mathscr{S}^t) = \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\eta^2}\right)\mathscr{S}_t$ and therefore,

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|\mathscr{S}^t\right)\right)^2\right] = \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_\varepsilon^2+\sigma_\eta^2} \tag{38}$$

Calculating the variance of the one-step-ahead forecast error for $s_t = (L-\vartheta)\varepsilon_t$ requires more careful treatment. The fundamental representation is derived through the use of Blaschke factors

$$s_t = (L-\vartheta)\left(\frac{1-\vartheta L}{L-\vartheta}\right)\left(\frac{L-\vartheta}{1-\vartheta L}\right)\varepsilon_t = (1-\vartheta L)e_t \tag{39}$$

$$e_t = \left(\frac{L-\vartheta}{1-\vartheta L}\right)\varepsilon_t \tag{40}$$

Given that (39) is an invertible representation then the Hilbert space spanned by current and past $x_t$ is equivalent to the space spanned by current and past $e_t$. This implies

$$\mathbb{E}(\varepsilon_t|e^t) = \mathbb{E}(\varepsilon_t|s^t) \tag{41}$$

To show (41) notice that (40) can be written as

$$\varepsilon_t = C(L)e_t = \left[\frac{1-\vartheta L}{L-\vartheta}\right]e_t = \left[\frac{L^{-1}-\vartheta}{1-\vartheta L^{-1}}\right]e_t = (L^{-1}-\vartheta)\sum_{j=0}^{\infty}\vartheta^j e_{t+j} \tag{42}$$

Thus, while (39) does not have an invertible representation in current and past $e$ it does have a valid expansion in current and future $e$. Applying the optimal prediction formula,

$$\mathbb{E}(\varepsilon_t|e^t) = \left[C(L)\right]_+ e_t = -\vartheta e_t = \left(\frac{-\vartheta}{1-\vartheta L}\right)s_t = \mathbb{E}(\varepsilon_t|s^t) \tag{43}$$

We must now calculate

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|s^t\right)\right)^2\right] = \mathbb{E}\left(\varepsilon_t^2\right) + \mathbb{E}\left(\varepsilon_t|s^t\right)^2 - 2\mathbb{E}\left(\varepsilon_t\mathbb{E}\left(\varepsilon_t|s^t\right)\right) \tag{44}$$

$$= \sigma_\varepsilon^2 + \vartheta^2\sigma_\varepsilon^2 - 2\mathbb{E}(\varepsilon_t\left(\varepsilon_t|s^t\right)) \tag{45}$$

where we've used the fact that the squared modulo of the Blaschke factor is equal to 1, $\left(\frac{1+\vartheta z}{z+\vartheta}\right)\left(\frac{1+\vartheta z^{-1}}{z^{-1}+\vartheta}\right) = 1$, and therefore $\mathbb{E}(e^2) = \sigma_\varepsilon^2$. To calculate $\mathbb{E}(\varepsilon_t\left(\varepsilon_t|s^t\right))$ we use complex integration and the theory of the residue calculus,

$$\mathbb{E}(\varepsilon_t e_t) = \frac{-\vartheta\sigma_\varepsilon^2}{2\pi i}\oint \frac{z-\vartheta}{1-\vartheta z}\frac{dz}{z} = \sigma_\varepsilon^2\left[\lim_{z\to 0}\frac{z-\vartheta}{1-\vartheta z}\right] = \vartheta^2\sigma_\varepsilon^2 \tag{46}$$

Equations (45) and (46) give the desired result

$$\mathbb{E}\left[\left(\varepsilon_t - E\left(\varepsilon_t|x^t\right)\right)^2\right] = \left(1-\vartheta^2\right)\sigma_\varepsilon^2 \tag{47}$$

Equating (47) and (38) concludes the proof,

$$\vartheta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\eta^2} \tag{48}$$

**A.2 PROOF OF PROPOSITION 1** The conditional expectations for the informed and uninformed are given by

$$\mathbb{E}_t^I(y_{t+1}) = L^{-1}[(L-\lambda)Y(L)+\lambda Y_0]\varepsilon_t$$
$$\mathbb{E}_t^U(y_{t+1}) = L^{-1}[(L-\lambda)Y(L)-Y_0\mathscr{B}_\lambda(L)]\varepsilon_t$$

Substituting the expectations into the equilibrium gives the $z$-transform in $\varepsilon_t$ space as

$$(z-\lambda)Y(z) = \beta\mu z^{-1}[(z-\lambda)Y(z)+\lambda Y_0] + \beta(1-\mu)z^{-1}[(z-\lambda)Y(z)-Y_0\mathscr{B}_\lambda(z)] + A(z)$$

and re-arranging yields the following functional equation

$$(z-\lambda)(z-\beta)Y(z) = zA(z) + \beta Y_0[\mu\lambda-(1-\mu)\mathscr{B}_\lambda(z)]$$

The $Y(\cdot)$ process will not be analytic unless the process vanishes at the poles $z = \{\lambda,\beta\}$. Evaluating at $z=\lambda$ gives the restriction on $A(\cdot)$, $A(\lambda) = -\beta\mu Y_0$. Rearranging terms

$$\begin{aligned}(z-\beta)Y(z) &= \frac{1}{z-\lambda}\left\{zA(z)+\beta Y_0[\mu\lambda-(1-\mu)\mathscr{B}_\lambda(z)]\right\}\\ &= \frac{1}{z-\lambda}\left\{zA(z)+\beta Y_0 h(z)\right\}\end{aligned} \tag{49}$$

where $h(z) \equiv [\mu\lambda-(1-\mu)\mathscr{B}_\lambda(z)]$. Evaluating at $z=\beta$ gives $Y_0 = -\frac{A(\beta)}{h(\beta)}$ to ensure stability. This implies that the restriction on $A(\cdot)$ is

$$A(\lambda) = \frac{\beta\mu A(\beta)}{h(\beta)}$$

which is (12). Substituting this into (49) delivers (13).

A.3 PROOF OF PROPOSITION 2 Similar to solving the previous model, the first step in the proof of Proposition 2 is to obtain an innovations representation for the signal vector $(\varepsilon_{it}, y_t)$ that can be used to formulate the expectation at the agent's level. That is, we must find the space spanned by current and past observables, $\{\varepsilon_{i,t-j}, y_{t-j}\}_{j=0}^{\infty}$. This representation in terms of the innovation $\varepsilon_t$ and the noise $v_{it}$ is

$$\begin{pmatrix} \varepsilon_{it} \\ y_t \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon & \sigma_v \\ (L-\lambda)\, Y(L) & 0 \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_t \\ \hat{v}_{it} \end{pmatrix} = \Gamma(L) \begin{pmatrix} \hat{\varepsilon}_t \\ \hat{v}_{it} \end{pmatrix} \tag{50}$$

where we have re-scaled the mapping so that the innovations $\hat{\varepsilon}_t$ and the noise $\hat{v}_{it}$ have unit variance. Let the fundamental representation be denoted by

$$\begin{pmatrix} \varepsilon_{it} \\ y_t \end{pmatrix} = \Gamma^*(L) \begin{pmatrix} w_{it}^1 \\ w_{it}^2 \end{pmatrix} \tag{51}$$

As with the hierarchical case, we must use Blaschke factors to flip the non-fundamental root, $\lambda$, to outside the unit circle. However, we must also employ a Gram-Schmidit type orthogonalization $(W_\lambda)$ so that the Blaschke factor does not introduce additional unstable roots into the dynamic process. This decomposition is given by

$$W_\lambda = \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_v^2}} \begin{pmatrix} \sigma_\varepsilon & -\sigma_v \\ \sigma_v & \sigma_\varepsilon \end{pmatrix}, \quad \mathscr{B}_\lambda(L) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1-\lambda L}{L-\lambda} \end{pmatrix}$$

$$\Gamma^*(L) = \Gamma(L) W_\lambda B_\lambda(L)$$

with the vector of fundamental innovations

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

The expectation term for agent $i$ is found by applying the the Wiener-Kolmogorov prediction formula to the fundamental representation (51)

$$\mathbb{E}(y_{t+1} | \varepsilon_i^t, y^t) = \left[\Gamma_{21}^*(L) - \Gamma_{21}^*(0)\right] L^{-1} w_{it}^1 + \left[\Gamma_{22}^*(L) - \Gamma_{22}^*(0)\right] L^{-1} w_{it}^2. \tag{52}$$

It is straightforward to show that

$$\begin{aligned} \Gamma_{21}^*(L) &= (L-\lambda)\, Y(L) \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}, \quad \Gamma_{21}^*(0) = -\lambda Y_0 \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}} \\ \Gamma_{22}^*(L) &= -(1-\lambda L)\, Y(L) \frac{\sigma_v}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}, \quad \Gamma_{22}^*(0) = -Y_0 \frac{\sigma_v}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}} \end{aligned}$$

Solving for the equilibrium requires averaging across all the agents. In taking those averages, the idiosyncratic components of the innovation (the noise) will be zero and one will have two terms that are functions only of the aggregate innovation, namely

$$\int_0^1 w_{it}^1 di = w_t^1 = \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}\hat{\varepsilon}_t \quad \text{and} \quad \int_0^1 w_{it}^2 di = w_t^2 = -\frac{\sigma_v}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}\frac{L-\lambda}{1-\lambda L}\hat{\varepsilon}_t.$$

The average market expectation is then

$$\bar{\mathbb{E}}(y_{t+1}) = [(L-\lambda)Y(L)+\lambda Y_0]\,L^{-1}\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2}\hat{\varepsilon}_t + [(1-\lambda L)Y(L)-Y_0]\,L^{-1}\frac{\sigma_v^2}{\sigma_\varepsilon^2+\sigma_v^2}\frac{L-\lambda}{1-\lambda L}\hat{\varepsilon}_t \tag{53}$$

Now, if we let

$$\tau \equiv \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2},$$

and substitute the functional form of the average expectations into the equilibrium equation for $y_t$, we would get

$$(L-\lambda)Y(L) = \beta\mu L^{-1}[(L-\lambda)Y(L)+\lambda Y_0] + \beta(1-\mu)L^{-1}\left[(L-\lambda)Y(L)+Y_0\frac{\lambda-L}{1-\lambda L}\right] + A(L)\sigma_\varepsilon$$

Setting $Y(L) = Q(L)\sigma_\varepsilon$, we can write the $z$-transform in $\varepsilon_t$ space of the fixed point condition

$$(z-\lambda)Q(z) = \beta\tau z^{-1}[(z-\lambda)Q(z)+\lambda Q_0] + \beta(1-\tau)z^{-1}\left[(z-\lambda)Q(z)+Q_0\frac{\lambda-L}{1-\lambda L}\right] + A(z) \tag{54}$$

Re-arranging yields the following functional equation

$$(z-\lambda)(z-\beta)Q(z) = zA(z) + \beta Q_0\left[\tau\lambda + (1-\tau)\frac{\lambda-z}{1-\lambda z}\right]$$

The $Q(\cdot)$ process will not be analytic unless the process vanishes at the poles $z = \{\lambda, \beta\}$. Evaluating at $z = \lambda$ gives the restriction on $A(\cdot)$, $A(\lambda) = -\beta\tau Q_0$. Rearranging terms

$$\begin{aligned}(z-\beta)Q(z) &= \frac{1}{z-\lambda}\left[zA(z)+\beta Q_0\left(\tau\lambda+(1-\tau)\frac{\lambda-z}{1-\lambda z}\right)\right]\\ &= \frac{1}{z-\lambda}\left[zA(z)+\beta Q_0 h(z)\right]\end{aligned} \tag{55}$$

where $h(z) \equiv \tau\lambda + (1-\tau)\frac{\lambda-z}{1-\lambda z}$. Evaluating at $z = \beta$ gives $Q_0 = -\frac{A(\beta)}{h(\beta)}$ to ensure stability; this also results in uniqueness. The fixed point for $\lambda$ can be then written as

$$A(\lambda) = \frac{\beta\mu A(\beta)}{h(\beta)}$$

which is (24). Substituting this into (55) delivers (25), which completes the proof.

**A.4 PROOF OF PROPOSITION 3** Once the analytic form for $\Gamma_{21}^*(L)$ and $\Gamma_{22}^*(L)$ are known from Proposition 2, one can compute $\mathbb{E}(y_{t+j}|\varepsilon_i^t, y^t)$ for any $j = 1, 2, \ldots$. We show the $j = 1$ case here. Substitute $\Gamma_{21}^*(L)$

and $\Gamma^*_{22}(L)$ into (52) and collecting the terms that constitute (53), one gets

$$\begin{aligned}
\mathbb{E}(y_{t+1}|\varepsilon_i^t, y^t) &= \bar{\mathbb{E}}(y_{t+1}) + \frac{\sigma_\varepsilon}{\sigma_\varepsilon^2+\sigma_v^2}L^{-1}[(L-\lambda)Y(L)+\lambda Y_0-(L-\lambda)Y(L)+Y_0\frac{L-\lambda}{1-\lambda L}]\sigma_v\hat{v}_{it} \\
&= \bar{\mathbb{E}}(y_{t+1}) + \frac{\sigma_\varepsilon}{\sigma_\varepsilon^2+\sigma_v^2}L^{-1}[\lambda Y_0+Y_0\frac{L-\lambda}{1-\lambda L}]\sigma_v\hat{v}_{it} \\
&= \bar{\mathbb{E}}(y_{t+1}) + \mu Y_0\frac{1-\lambda^2}{1-\lambda L}v_{it},
\end{aligned} \tag{56}$$

which completes the proof for the first statement of the theorem for $j = 1$. The variance of the term $\mu Y_0 \frac{1-\lambda^2}{1-\lambda L} v_{it}$ can be readily computed since the innovations $v_{it}$ are independently distributed with variance $\sigma_v^2$.

**A.5 HOBS WITH HIERARCHICAL INFORMATION** Write the equilibrium as $y_t = (L-\lambda)Y(L)\varepsilon_t$ where $|\lambda|<1$ and $Y(L)$ satisfies Proposition 1. For $j = 1$, the time $t+1$ average expectation at $t+2$ is given by

$$\begin{aligned}
\overline{\mathbb{E}}_{t+1}y_{t+2} &= \mu\mathbb{E}^I_{t+1}y_{t+2} + (1-\mu)\mathbb{E}^U_{t+1}y_{t+2} \\
&= L^{-1}(L-\lambda)Y(L)\varepsilon_{t+1} + L^{-1}Y_0[\mu\lambda-(1-\mu)\mathscr{B}_\lambda(L)]\varepsilon_{t+1} \\
&= y_{t+2} + L^{-1}Y_0[\mu\lambda-(1-\mu)\mathscr{B}_\lambda(L)]\varepsilon_{t+1}
\end{aligned} \tag{57}$$

The informed agent's time $t$ expectation of the average expectation at $t+1$ is

$$\mathbb{E}^I_t\overline{\mathbb{E}}_{t+1}y_{t+2} = \mathbb{E}^I_t y_{t+2} + \mu\lambda Y_0\mathbb{E}^I_t\varepsilon_{t+2} - Y_0(1-\mu)\mathbb{E}^I_t\mathscr{B}_\lambda(L)\varepsilon_{t+2}. \tag{58}$$

Clearly $\mathbb{E}^I_t\varepsilon_{t+2} = 0$, whereas the expectation in the last term of (58) is given by

$$\mathbb{E}^I_t\mathscr{B}_\lambda(L)\varepsilon_{t+2} = L^{-2}\{\mathscr{B}_\lambda(L)-\mathscr{B}_\lambda(0)-\mathscr{B}_\lambda(1)L\}\varepsilon_t \tag{59}$$

where the notation $\mathscr{B}_\lambda(j)$ stands for "the sum of the coefficients of $L^j$". If we write

$$\mathscr{B}_\lambda(L) = (L-\lambda)(1+\lambda L+\lambda^2L^2+\lambda^3L^3+\cdots).$$

it is straightforward to show that $\mathscr{B}_\lambda(0) = -\lambda$ and $\mathscr{B}_\lambda(1) = (1-\lambda)(1+\lambda) = (1-\lambda^2)$, from which follows

$$\mathscr{B}_\lambda(L)-\mathscr{B}_\lambda(0)-\mathscr{B}_\lambda(1)L = \frac{L-\lambda}{1-\lambda L}+\lambda-(1-\lambda^2)L = \frac{\lambda(1-\lambda^2)L^2}{1-\lambda L}.$$

Putting things together, the informed agent's expectation of the average expectation is

$$\mathbb{E}^I_t\overline{\mathbb{E}}_{t+1}y_{t+2} = \mathbb{E}^I_t y_{t+2} - (1-\mu)Y_0\lambda\Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_t \tag{60}$$

For the uninformed,

$$\mathbb{E}^U_t\overline{\mathbb{E}}_{t+1}y_{t+2} = \mathbb{E}^U_t y_{t+2} + Y_0\mu\lambda\mathbb{E}^U_t\varepsilon_{t+2} - Y_0(1-\mu)\mathbb{E}^U_t\mathscr{B}_\lambda(L)\varepsilon_{t+2}$$

As for the informed case, $\mathbb{E}_t^U \varepsilon_{t+2} = 0$; however, the second term now is $\mathbb{E}_t^U \mathscr{B}_\lambda(L)\varepsilon_{t+2} = 0$ because, by definition, $\mathscr{B}_\lambda(L)\varepsilon_{t+2}$ is not in the information set of the uninformed agents at time $t$. Hence $\mathbb{E}_t^U \overline{\mathbb{E}}_{t+1} y_{t+2} = \mathbb{E}_t^U y_{t+2}$: the uninformed are *not* forming higher-order expectations.
Applying the above results to the market forecast of the market forecast one gets

$$\overline{\mathbb{E}}_t \overline{\mathbb{E}}_{t+1} y_{t+2} = \mu \mathbb{E}_t^I \overline{\mathbb{E}}_{t+1} y_{t+2} + (1-\mu)\mathbb{E}_t^U \overline{\mathbb{E}}_{t+1} y_{t+2} \qquad = \overline{\mathbb{E}}_t y_{t+2} - \mu(1-\mu) Y_0 \lambda \Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_t, \tag{61}$$

which shows that the market forecast operator does not satisfy the law of iterated mathematical expectations. We can now characterize the entire structure of the market HOB. For $j = 2$, we need to calculate $\overline{\mathbb{E}}_t \overline{\mathbb{E}}_{t+1} \overline{\mathbb{E}}_{t+2} y_{t+3}$. From (57),

$$\overline{\mathbb{E}}_{t+2} y_{t+3} = y_{t+3} + Y_0[\mu\lambda - (1-\mu)\mathscr{B}_\lambda(L)]\varepsilon_{t+3}$$

We then need the uninformed and informed's time $t+1$ expectations of $\overline{\mathbb{E}}_{t+2} y_{t+3}$. For the uninformed we know from above (taking the time one period forward) that $\mathbb{E}_{t+1}^U \overline{\mathbb{E}}_{t+2} y_{t+3} = \mathbb{E}_{t+1}^U y_{t+3}$. From standard conditioning down one has

$$\begin{aligned}\mathbb{E}_{t+1}^U y_{t+3} &= \Big[\frac{(1-\lambda L) Y(L)}{L^2}\Big] + \mathscr{B}_\lambda(L)\varepsilon_{t+1} \\ &= L^{-2}[(L-\lambda) Y(L) - (Y_0 + (Y_1 - \lambda Y_0)L)\mathscr{B}_\lambda(L)]\varepsilon_{t+1} \end{aligned} \tag{62}$$

For the informed

$$\begin{aligned}\mathbb{E}_{t+1}^I \overline{\mathbb{E}}_{t+2} y_{t+3} &= \mathbb{E}_{t+1}^I y_{t+3} + \mu Y_0 \lambda \mathbb{E}_{t+1}^I \varepsilon_{t+3} - (1-\mu) Y_0 \mathbb{E}_{t+1}^I \mathscr{B}_\lambda(L)\varepsilon_{t+3} \\ &= L^{-2}[(L-\lambda) Y(L) + \lambda Y_0 - (Y_0 - \lambda Y_1)L]\varepsilon_{t+1} - (1-\mu) Y_0 \lambda \Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_{t+1}. \end{aligned} \tag{63}$$

Combining (62) and (63) gives

$$\begin{aligned}\overline{\mathbb{E}}_{t+1} \overline{\mathbb{E}}_{t+2} y_{t+3} = y_{t+3} &+ \mu\{\lambda Y_0 - (Y_0 - \lambda Y_1)L\}\varepsilon_{t+3} - \mu(1-\mu) Y_0 \lambda \Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_{t+1} \\ &-(1-\mu)[Y_0 + (Y_1 - \lambda Y_0)L]\mathscr{B}_\lambda(L)\varepsilon_{t+3} \end{aligned} \tag{64}$$

Following the same argument that we used for the first order expectations it is easy to conclude that the uninformed's expectations of (64) are just

$$\mathbb{E}_t^U \overline{\mathbb{E}}_{t+1} \overline{\mathbb{E}}_{t+2} y_{t+3} = \mathbb{E}_t^U y_{t+3} \tag{65}$$

This is because the uninformed cannot forecast the informed forecast of their forecast error; for the uninformed such forecast error belongs to information they will only receive in the future. Formally

$$\mathbb{E}_t^U \Big(\frac{1}{1-\lambda L}\Big)\varepsilon_{t+1} = \mathbb{E}_t^U \Big(\frac{1}{L-\lambda}\Big) e_{t+1} = \mathbb{E}_t^U \sum_{j=0}^{\infty} \lambda^j e_{t+1} = 0.$$

For the informed

$$\mathbb{E}_t^I\overline{\mathbb{E}}_{t+1}\overline{\mathbb{E}}_{t+2}y_{t+3} = \mathbb{E}_t^I y_{t+3} - Y_0\mu(1-\mu)\lambda^2\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t - Y_1(1-\mu)\lambda\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t$$
$$= \mathbb{E}_t^I y_{t+3} - (1-\mu)(Y_0\mu\lambda^2 + Y_1\lambda)\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t$$

Therefore the average expectation is

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\overline{\mathbb{E}}_{t+2}y_{t+3} = \overline{\mathbb{E}}_t y_{t+3} - (1-\mu)(Y_0\mu^2\lambda^2 + Y_1\mu\lambda)\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t. \tag{66}$$

Comparing this to (61) one can already see a pattern in the coefficients multiplying the noise term related to the forecast error of the uninformed. Iterating the process over and over one obtains the generic form of the higher order market expectations for prices

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\cdots\overline{\mathbb{E}}_{t+j}y_{t+j+1} = \overline{\mathbb{E}}_t y_{t+j+1} - (1-\mu)\left(\sum_{i=1}^{j}(\mu\lambda)^i Y_{j-i}\right)\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t$$

### A.6 HoBs: Dispersed Information Case

We begin by noticing that

$$\mathbb{E}_{it}\overline{\mathbb{E}}_{t+1}y_{t+2} = \mu\mathbb{E}_{it}\mathbb{E}_{t+1}^I y_{t+2} + (1-\mu)\mathbb{E}_{it}\mathbb{E}_{t+1}^U y_{t+2}. \tag{67}$$

From the hierarchical equilibrium, we know that $\mathbb{E}_{t+1}^U y_{t+2} = \mathbb{E}_{t+1}^I y_{t+2} - Y_0\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}$. We also notice that, because the information set of an arbitrary agent $i$ is strictly smaller than the information set of an informed agent of the hierarchical equilibrium and because the law of iterated expectations holds at the single agent level, we have $\mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}_{t+1}^I y_{t+2} = \mathbb{E}_{it}y_{t+2}$. Because of the second property we also have that $\mathbb{E}_{it}\mathbb{E}_{t+1}^U y_{t+2} = \mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}_{t+1}^U y_{t+2}$. Therefore

$$\mathbb{E}_{it}\overline{\mathbb{E}}_{t+1}y_{t+2} = \mu\mathbb{E}_{it}y_{t+2} + (1-\mu)\mathbb{E}_{it}y_{t+2} - (1-\mu)Y_0\mathbb{E}_{it}\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}. \tag{68}$$

The crucial step in the proof is then to show that the expectation in the last term is non-zero. In order to do so we first notice that $\frac{L-\lambda}{1-\lambda L}\varepsilon_{t+2} = \frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1} - \lambda\varepsilon_{t+2}$ and so

$$\mathbb{E}\left(\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}|\varepsilon_i^t, y^t\right) = \mathbb{E}\left(\frac{L-\lambda}{1-\lambda L}\varepsilon_{t+2}|\varepsilon_i^t, y^t\right). \tag{69}$$

Then, the crucial step in the proof is to show that

$$\mathbb{E}\left(\frac{L-\lambda}{1-\lambda L}\varepsilon_{t+2}|\varepsilon_i^t, y^t\right) = \mu\lambda\frac{(1-\lambda^2)}{1-\lambda L}\varepsilon_{it}. \tag{70}$$

where $\mu \equiv \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2}$. Our prediction formula follows that of Theorem 1 and Whittle (1983)

$$\left[L^{-2} g_{e,(\varepsilon,y)}(L)\left(\Gamma^*(L^{-1})^T\right)^{-1}\right]_+ \Gamma^*(L)^{-1} \tag{71}$$

where $\Gamma^*(L)$ is defined in (51) and $g_{e,(\varepsilon,y)}(L)$ is the variance-covariance generating function between the variable to be predicted and the variables in the information set. In our case we have that

$$g_{e,(\varepsilon,y)}(L) = \left[\mathscr{B}(L)\sigma_\varepsilon^2 \quad \mathscr{B}(L)\left(L^{-1}-\lambda\right)Y\left(L^{-1}\right)\sigma_\varepsilon\right]$$

Plugging in the explicit forms and solving out the algebra

$$L^{-2} g_{e,(\varepsilon,y)}(L)\left(\Gamma^*(L^{-1})^T\right)^{-1} = \frac{1}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}\left[L^{-2}\frac{L-\lambda}{1-\lambda L}\sigma_\varepsilon^2 + L^{-2}\left(L^{-1}-\lambda\right)Y\left(L^{-1}\right)\frac{\sigma_\varepsilon^2}{\sigma_v} \quad -L^{-2}\frac{\sigma_\varepsilon^2+\sigma_v^2}{\sigma_v}\sigma_\varepsilon\right]$$

Applying the annihilator operator to the RHS we see that the second term of the vector goes to zero. For the first term, the assumption that $Y(L)$ is analytic inside the unit circle ensures that $L^{-2}\left(L^{-1}-\lambda\right)Y\left(L^{-1}\right)$ does not contain any term in positive power of $L$. We are then left with

$$\left[L^{-2}\frac{L-\lambda}{1-\lambda L}\right]_+ = \frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}, \tag{72}$$

Summarizing we have shown that

$$\mathbb{E}_{it}\left(\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}\right) = \frac{1}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}\frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}\sigma_\varepsilon^2 w_{it}^1$$

Substituting in $w_{it}^1 = \frac{1}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}(\varepsilon_t + v_{it})$ completes the proof. The proof can be generalized to expectations of order higher than 1 following the same pattern as the derivations in the hierarchical case.

A.7 PROOF OF COROLLARY 3 The proof follows immediately from the restriction (12). Condition (4.a) is derived by taking the limit of (12) as $\mu \to 0$ (or equivalently $\tau \to 0$). This is the equilibrium that would exist if no informed agents populated the model. Intuitively, if no hierarchical information equilibrium exists in this case, then none would exist if informed agents had positive measure. This restriction is given by $A(\lambda) = 0$ for $|\lambda| < 1$, which for the process $A(\lambda) = (1+\theta\lambda)/(1-\rho\lambda)$, implies $\theta \in (0,1)$. Notice that because $\theta > 0$, $\lambda \to -1$ from above. Substituting $\lambda = -1$ into (12) and solving for $\mu$ gives condition (4.c). When $\lambda = -1$, the equilibrium converges to the full-information case. Setting $\mu^\star$ equal to unity and solving for $\theta$ gives condition (4.b).

# B HOMOGENEOUS-BELIEFS ECONOMIES

The model consists of an equilibrium equation and a stochastic, exogenous process

$$y_t = \beta \mathbb{E}_t y_{t+1} + x_t \tag{73}$$

$$x_t = A(L)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{74}$$

where $x_t = A(L)\varepsilon_t = A_0\varepsilon_t + A_1\varepsilon_{t-1} + \cdots$, $L$ is a lag operator $Lx_t \equiv x_{t-1}$, and the coefficients satisfy square summability, $\sum_j A_j^2 < \infty$. Representation (74) places no restrictions on the serial correlation properties of $x_t$. The Wold Decomposition Theorem allows for such a general representation.

## B.1 FULL INFORMATION

Following standard procedure, we look for a solution of the endogenous variable, $y_t$, that satisfies square summability and exists in the agents' information set. The full-information solution assumes that the agents have perfect knowledge of current and past shocks. Denote this full information as $FI$ with an information set, $\Omega_t^{FI} = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$, which suggests a guess for the equilibrium of the form $y_t = Y(L)\varepsilon_t = Y_0\varepsilon_t + Y_1\varepsilon_{t-1} + \cdots$. Conditional expectations are evaluated via the Wiener-Kolmogorov optimal prediction formula,

$$\begin{aligned}\mathbb{E}_t^{FI}[y_{t+1}] &= \mathbb{E}[Y(L)\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, ...] = L^{-1}[Y(L) - Y_0]\varepsilon_t \\ &= L^{-1}[Y_0 + Y_1 L + Y_2 L + \cdots - Y_0]\varepsilon_t = Y_1\varepsilon_t + Y_2\varepsilon_{t-1} + \cdots \end{aligned} \tag{75}$$

The prediction formula instructs the agent to subtract off the $\varepsilon_{t+1}$ term as it does not enter the agents' information set and has an expected value of zero.

Substituting the equilibrium guess $y_t = Y(L)\varepsilon_t$ and the expectation (75) into equation (73) gives $Y(L)\varepsilon_t = \beta L^{-1}[Y(L) - Y_0]\varepsilon_t + A(L)\varepsilon_t$. We use techniques first established in Whiteman (1983) that rely on analytic function theory to solve for the rational expectations equilibrium. This methodology invokes the Riesz-Fischer Theorem, which states that the sequential problem of finding $Y_0, Y_1, Y_2, ...$ has an equivalent representation as a functional problem in the Hardy space of analytic functions $Y(z)$. Our problem becomes one of finding the function $Y(z)$ that solves

$$\begin{aligned} Y(z) &= \beta z^{-1}[Y(z) - Y_0] + A(z) \\ &= \frac{zA(z) - Y_0}{z - \beta} \end{aligned} \tag{76}$$

Following a long tradition in rational expectation modeling, we look for solutions to the sequential problem that satisfy square summability, $\sum_j Y_j^2 < \infty$ (i.e., we look for bounded or stationary equilibria). Square summability is tantamount to analyticity inside the unit circle in the space of $z$-transforms. The $Y(z)$ process given by (76) has a pole at $z = \beta$. If $|\beta| > 1$, the $Y(z)$ process is analytic inside the unit circle but has a undetermined parameter $Y_0$. In this case, $Y_0$ cannot be uniquely pinned down and the rational expectations model has an infinite number of equilibria. If $|\beta| < 1$, the process is not analytic inside the unit circle and $Y_0$ is needed to remove the pole at $z = \beta$, which gives $Y_0 = \beta A(\beta)$. Under this scenario, the rational expectations solution is unique and given by

$$Y(z) = \frac{zA(z) - \beta A(\beta)}{z - \beta} \tag{77}$$

which is the ubiquitous Hansen-Sargent formula [Hansen and Sargent (1980)].

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 that we take advantage of throughout the paper. The first component, $zA(z)/(z-\beta)$, is the perfect foresight equilibrium; that is, iterate (73) forward, impose the law of iterated expectations and a no-bubble condition to solve

$$y_t = \mathbb{E}_t^{FI} \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t^{FI} \left( \frac{LA(L)}{L-\beta} \right) \varepsilon_t \tag{78}$$

If we appended the agents' information set with future values of $\varepsilon_t$, such that agents have perfect foresight (PF) $\Omega_t^{PF} = \{\varepsilon_{t-j}\}_{j=-\infty}^{\infty}$, (78) (after removing the expectation operator) would be the rational expectations equilibrium. Therefore the last element of the H-S formula, $\beta A(\beta)/(z-\beta)$, represents the conditioning down associated with only observing current and past $\varepsilon_t$'s. Subtracting off this precise linear combination of future shocks, $\beta A(\beta) \sum_j \beta^j \varepsilon_{t+j}$, stems from knowledge that the model is given by (73)-(3) and the information set of $\Omega_t^{FI} = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$.[7]

B.2 INCOMPLETE INFORMATION Working within a representative agent framework, we now derive an equilibrium with incomplete information. By incomplete information, we mean an equilibrium that exists in a subspace of the sequence generated by the fundamental shocks, $\{\varepsilon_{t-j}\}_{j=0}^{\infty}$.

Returning to our endogenous signal extraction problem of (4), we must first find the corresponding innovations associated with observing current and past $y_t$; thus, we must flip the $\lambda$ root from inside the unit circle to outside the unit circle without changing the moments of the $y_t$ process. This transformation is accomplished through the use of Blaschke factors, $\mathscr{B}_\lambda(L) \equiv (L-\lambda)/(1-\lambda L)$

$$y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t = (1-\lambda L)\widetilde{Y}(L)e_t \tag{79}$$

$$e_t = \left( \frac{L-\lambda}{1-\lambda L} \right) \varepsilon_t = (L-\lambda)(\varepsilon_t + \lambda \varepsilon_{t-1} + \lambda^2 \varepsilon_{t-1} + \cdots) \tag{80}$$

Note that we are operating in well-defined Hilbert spaces with the covariance generating function serving as the modulus and that Blaschke factors have a modulus of one, $\mathscr{B}_\lambda(z)\mathscr{B}_\lambda(z^{-1}) = 1$, supporting the equality in (79). Note also that *conditional* expectations differ in the $e_t$ and $\varepsilon_t$ spaces.

The guess of the equilibrium process (79) must be verified, and uniquely so. This is accomplished by forming the expectation conditional on Partial Information (PI)

$$\mathbb{E}[y_{t+1} | \Omega_t^{PI} = \{y_{t-j}\}_{j=0}^{\infty}] = \mathbb{E}[(1-\lambda L)\widetilde{Y}(L)e_{t+1}] = L^{-1}[(1-\lambda L)\widetilde{Y}(L) - \widetilde{Y}_0]e_t \tag{81}$$

[7] As shown in Appendix A of Hansen and Sargent (1980), agents who know the model is given by (78) will form expectations optimally by subtracting off the principal part of the Laurent series expansion of $A(z)$ around $\beta$, which is $\beta A(\beta)/(z-\beta)$.

which is simply the Wiener-Kolmogorov optimal prediction formula applied to (79). Substituting this expectation into (73) gives

$$(1-\lambda L)\widetilde{Y}(L)\mathscr{B}_\lambda(L)\varepsilon_t = \beta L^{-1}[(1-\lambda L)\widetilde{Y}(L)-\widetilde{Y}_0]\mathscr{B}_\lambda(L)\varepsilon_t + A(L)\varepsilon_t$$

We then repeat the functional analysis described above by solving for $y_t$, assuming $\beta \in (0,1)$,

$$(z-\lambda)\widetilde{Y}(z) = \frac{zA(z)-\beta A(\beta)\mathscr{B}_\lambda(z)/\mathscr{B}_\lambda(\beta)}{z-\beta} \tag{82}$$

However, there is an additional step that we must take in order to prove that the expectation is consistent with (81) and that the sequence $\{y_{t-j}\}_{j=0}^{\infty}$ does *not* reveal $\varepsilon_t$. We assumed that the endogenous variable is not invertible in $\lambda$, this is only true if the RHS of (82) vanishes at $z=\lambda$. We have also assumed that there is only one zero inside the unit circle (i.e., $\widetilde{Y}(L)$ contains no zeroes inside the unit circle). This places restriction on the exogenous process, namely, $A(\lambda)=0$, which we write as $x_t=(L-\lambda)\widetilde{A}(L)\varepsilon_t$, where $\widetilde{A}(L)$ does not have any zeros inside the unit circle. If this restriction holds (which is tantamount to assuming the exogenous process is not fundamental for $\varepsilon_t$), then the unique rational expectations equilibrium is given by (82). If this restriction does not hold, then the endogenous variable will completely reveal the underlying shocks and the equilibrium will be the full-information equilibrium of Section B.1. We have proved the following:

**Proposition 4.** *Consider the economy described by* (73)–(74) *with expectations given by* $\mathbb{E}[y_{t+1}|\{y_{t-j}\}_{j=0}^{\infty}]$. *If* $\beta\in(0,1)$ *and*

$$A(\lambda)=0 \tag{83}$$

*with* $|\lambda|\in(0,1)$*, then the unique rational expectations equilibrium is given by*

$$y_t = \left(\frac{L(1-\lambda L)\widetilde{A}(L)-\beta(1-\lambda\beta)\widetilde{A}(\beta)}{L-\beta}\right)e_t \tag{84}$$
$$e_t = \left(\frac{L-\lambda}{1-\lambda L}\right)\varepsilon_t$$

*If* $|\lambda|>1$*, then the rational expectations equilibrium is unique and given by* (77).

From the perspective of the uninformed agents, the model lives in the $e_t$ space as shown by (84). The model is interpreted as solving the following discounted expectation,

$$y_t = \mathbb{E}_t^U\sum_{j=0}^{\infty}\beta^j x_{t+j} = \mathbb{E}_t^U\left(\frac{L(1-\lambda L)\widetilde{A}(L)}{L-\beta}\right)e_t \tag{85}$$

As with the full-information case, subtracting off the corresponding linear combination of future shocks, $\beta(1-\lambda\beta)\widetilde{A}(\beta)\sum_j\beta^j e_{t+j}$, delivers the conditioning down term of the rational expectations equilibrium in (84). However, the following corollary derives the equilibrium in the $\varepsilon_t$ space.

**Corollary 4.** *There is an equivalent representation of the equilibrium of Proposition 4 given by*

$$y_t = \left( \frac{L(L-\lambda)\widetilde{A}(L) - \beta(\beta-\lambda)\widetilde{A}(\beta)}{L-\beta} \right) \varepsilon_t - \left[ \frac{\beta \widetilde{A}(\beta)(1-\lambda^2)}{1-\lambda L} \right] \varepsilon_t \tag{86}$$

Representation (84) is the equilibrium in $e_t$ space and (86) is the equilibrium in $\varepsilon_t$ space. They are equivalent representations of the same equilibrium. Representation (84) is the standard looking Hansen-Sargent formula because this is the space that contains the agents' information set (current and past $e_t$'s). The first element on the right-hand side of (86) is the Hansen-Sargent formula under full information. The last term on the RHS represents the conditioning down due to partial information. Notice that as $|\lambda|$ approaches one from below, this term vanishes and the model converges to the full-information equilibrium.

To shed light on the representation (86), note the straightforward decomposition

$$\mathbb{E}_t^U \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t^I \left( \sum_{j=0}^{\infty} \beta^j x_{t+j} \right) - \beta \tilde{A}(\beta)(1-\lambda^2) \sum_{k=0}^{\infty} \lambda^k \varepsilon_{t-k} \tag{87}$$

The uninformed agents' expectations of fundamentals at each future date can be written as a linear combination of the expectation assuming agents see current and past structural shocks $\mathbb{E}_t^{FI}(x_{t+j})$, and a term given by linear combination of past $\varepsilon_t$'s that the agents do not observe. Notice that the linear combination is just the dynamic noise term of equation (6) multiplied by the weight $\beta\tilde{A}(\beta)$. As we show below, the representation of Corollary 4 is particularly useful when interpreting equilibrium objects like higher-order beliefs.