EQUIVALENCE RESULTS WITH ENDOGENOUS SIGNAL EXTRACTION*

Giacomo Rondina† Todd B. Walker‡

May 2023

Abstract

We derive equivalence results in dynamic models with information frictions to help solve for equilibrium and facilitate interpretation. Our primary theorem delivers an equivalence, in the aggregate, between 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 much easier to solve and interpret, especially when agents infer information from endogenous sources. We also generalize the ubiquitous Hansen-Sargent formula to models with incomplete information and derive equivalence-class representations as a function of information. 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.

Keywords: Endogenous Signal Extraction, Higher-Order Beliefs
JEL Codes: D82, D83

*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

Models with incomplete information offer a rich set of results unobtainable in representative agent, rational expectations economies that have implications for business cycle modeling, asset pricing and optimal policy, among others. And yet solving these models remains a challenge, especially when agents infer information from endogenous sources and information is dispersed evenly among all agents. We derive several "equivalence results" to facilitate solving and interpreting models in this environment.

We present a novel equivalence result that facilitates the characterization of equilibria in this environment. More precisely, we show that the aggregate representation of an equilibrium in models with dispersed information is isomorphic that of a model with two types of agents: informed and uninformed. We use the equivalence to gain important insights on three aspects of models of incomplete information. First, their relationship with the Hansen-Sargent formula, which imposes "cross-equations" restrictions important for empirical identification. Second, the characterization of higher-order beliefs, which escapes full formalization in the absence of an equilibrium solution. Third, the role of structural model parameters for the endogenous information transmission, which is typically ignored due to its complex characterization, but that can potentially constitute an important amplification/propagation channel.

Dispersed-Hierarchical Equivalence. Theorem 1 provides an equivalence, 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. 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.

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 signal in the dispersed information economy. This equivalence result can be understood by thinking about the optimal signal extraction strategy of dispersedly informed agents as a mixed strategy. With some probability, agents will act as if their private sig-

nal is exactly correct, mimicking the behavior of the perfectly informed agents. With the complementary probability, they will act as if their private signal contains no information about the state, mimicking the behavior of uninformed agents. While individual forecasts maintain a well defined cross-sectional distribution of beliefs (Proposition 4), the idiosyncratic noise component does not survive aggregation, which delivers our aggregate equivalence. The generality of this result extends beyond the setting discussed herein; Theorem 1 can be applied broadly to many models, even when analytical tractability is no longer feasible. Theorem 1 allows us to gain important insights into the interaction of information and dynamics in equilibrium which we exploit in analyzing higher-order beliefs.

Equivalent Hansen-Sargent Representations. Since the rational expectations revolution, solving for equilibria in dynamic models relies on imposing a mapping from exogenous stochastic processes to endogenous variables through optimal behavior. This mapping is often referred to as the Hansen-Sargent (1980) formula and makes clear the cross-equation restrictions, which are the "hallmark of rational expectations models," Sargent (1981). Every dynamic model equilibrium with an expectation operator (even those that are not rational) has a Hansen-Sargent representation.

Corollaries 1–3 generalize the Hansen-Sargent formula to models with information frictions and derive equivalent representations to facilitate interpretation. These formula make transparent the conditioning down necessary as information degrades from a perfect foresight to a full information to an incomplete information equilibria. In models with heterogeneous beliefs, the Hansen-Sargent formula reveals that each agent type views the "market fundamental" as a linear combination of the discounted sum of the exogenous process and the forecast *errors* of the other agent types. Thus, agents are forecasting the forecast errors of others, an addendum to the title of Townsend's (1983b) seminal work. This insight is the key to deriving restrictions on the exogenous processes that ensure incomplete information is preserved in equilibrium. When endogenous variables transmit information, the equilibrium fixed point problem typical of the rational expectations paradigm involves a mapping from endogenous variables to the agents' information set: given the equilibrium obtained under the expectations specified for a given information set, the information revealed in equilibrium should be consistent with the information used to solve for the equilibrium. In dynamic settings with incomplete information, this fixed point condition is nontrivial and a crucial aspect of the equilibrium. Propositions 1 and 2 contain two equations—one that describes the equlibrium dynamics and one that provides restrictions on exogenous processes that guarantee uninformed agents do not learn too much from endogenous signal extrac-

tion. The latter equation falls naturally from our generalized Hansen-Sargent formulas.

Higher-Order Beliefs. Equipped with an analytical characterization of the market equilibria under dispersed information due to Theorem 1 and equivalent Hansen-Sargent representations of Corollaries 1–3, 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.

In a dispersed information setting where every agent is equally uninformed, HoBs do not exist in the traditional form. Agent i does not forecast the forecasts of agent j. At the invidual level, each agent must think her information superior to that of other agents in order for HoBs of this type to be optimal. How and why are HoBs formed? The mixed-strategy intuition behind Theorem 1 provides the answer. 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 (i.e., agents inferring the signal as pure noise), 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.

Information Transmission. We 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 it changes with these parameters and higher-order belief dynamics. Corollary 4 derives a restriction on the informativeness of the signal in the dispersed information economy as a function of deep parameters. 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. That is, we 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.

Contacts with Literature. 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), and Han et al. (2023) employ frequency domain techniques to solve for a rational expectations equilibrium with some form of information friction. One distinguishing feature of this paper relative to many of those listed above is our emphasis on Hansen-Sargent (1980) formulas. The Hansen-Sargent formula naturally follows from these techniques and we provide an informational interpretation of this equation.

Some form of Theorem 1 is likely operational in dynamic models when information is dispersed according to a noisy signal of the underlying state. 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.

Equivalence results have been employed in the incomplete information literature to great effect over the years. For example, needing to find a way to compact a potentially infinite dimensional state space, Sargent (1991) first recognized that low-order ARMA processes could mimic infinite-dimensional moving average representations. Kasa (2000)

pushed this interpretation further by showing the ease with which these calculations are completed in the frequency domain. 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 MODELS AND EQUILIBRIUM REPRESENTATIONS

For transparency, we derive our equivalence results within the context of a generic, univariate rational expectations model and discuss extensions in Section 4. As a benchmark, we begin with a full-information, representative agent formulation. The model consists of an equilibrium equation and a stochastic, exogenous process

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

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

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 (2) places no restrictions on the serial correlation properties of x_t. The Wold Decomposition Theorem allows for such a general representation.

2.1 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 or "Informed" (I) information set as, $\Omega_t^I = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$, 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,

$$\begin{aligned}\mathbb{E}_t^I[y_{t+1}] &= \mathbb{E}[Y(L)\varepsilon_{t+1}|\varepsilon_t, \varepsilon_{t-1}, ...] = L^{-1}[Y(L) - Y_0]\varepsilon_t \\ &= L^{-1}[Y_0 + Y_1 L + Y_2 L + \cdots - Y_0]\varepsilon_t = Y_1\varepsilon_t + Y_2\varepsilon_{t-1} + \cdots\end{aligned} \tag{3}$$

The prediction formula instructs the agent to subtract off the ε_{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 (3) into equation (1) 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) and use 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

$$\begin{aligned} Y(z) &= \beta z^{-1}[Y(z) - Y_0] + A(z) \\ &= \frac{zA(z) - Y_0}{z - \beta} \end{aligned} \tag{4}$$

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 (4) 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

$$Y(z) = \frac{zA(z) - \beta A(\beta)}{z - \beta} \tag{5}$$

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 (1) forward, impose the law of iterated expectations and a no-bubble condition to solve

$$y_t = \mathbb{E}_t^I \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t^I \left(\frac{LA(L)}{L - \beta} \right) \varepsilon_t \tag{6}$$

If we appended the agents' information set with future values of ε_t, $\Omega_t^{PF} = \{\varepsilon_{t-j}\}_{j=-\infty}^{\infty}$, (6) (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 ε_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 (1)-(2) and the information set of $\Omega_t^{FI} = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$.[1]

[1] As shown in Appendix A of Hansen and Sargent (1980), agents who know the model is given by (6)

2.2 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 subset of the sequence generated by $\{\varepsilon_{t-j}\}_{j=0}^{\infty}$. One manner to derive such an equilibrium is to show that the endogenous process is given by[2]

$$y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t \qquad |\lambda| \in [0,1) \tag{7}$$

If $|\lambda| \in [0,1)$, 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 will be "Uninformed" (U) 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 ε_t sequence, and thus the information set generated by observing the y_t's is a strict subset of that generated by the ε_t's.

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

$$s_t = -\vartheta\varepsilon_t + \varepsilon_{t-1} = (L-\vartheta)\varepsilon_t, \tag{8}$$

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

$$\mathbb{E}\left(\varepsilon_t|\{s_{t-j}\}_{j=0}^{\infty}\right) = \underbrace{\vartheta^2\varepsilon_t}_{\text{information}} \quad - \quad \underbrace{(1-\vartheta^2)[\vartheta\varepsilon_{t-1}+\vartheta^2\varepsilon_{t-2}+\vartheta^3\varepsilon_{t-3}+\cdots]}_{+\ \text{noise from confounding dynamics}}. \tag{9}$$

Expression (9) suggests that the process (8) is informationally equivalent to a noisy signal about ε_t, where the noise is the linear combination of past shocks (in the bracketed term), and the signal-to-noise ratio is measured by ϑ^2. A ϑ 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 ε_t and the optimal prediction is 0, the unconditional average. As long as $|\vartheta| \in (-1,1)$, the value of ε_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 reve-

will form expectations optimally by subtracting off the principal part of the Laurent series expansion of $A(z)$ around β, which is $\beta A(\beta)/(z-\beta)$.

[2]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).

lation 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, Rondina and Walker (2021) show that the more standard 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

$$\mathscr{S}_t = \varepsilon_t + \eta_t, \tag{10}$$

where $\eta_t \overset{iid}{\sim} N\left(0, \sigma_\eta^2\right)$. The optimal prediction is well known and given by $\mathbb{E}(\varepsilon_t|\mathscr{S}^t) = \tau\mathscr{S}_t$, where τ 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,

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|s^t\right)\right)^2\right] = \mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|\mathscr{S}^t\right)\right)^2\right] \tag{11}$$

when

$$\vartheta^2 = \tau = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\eta^2} \tag{12}$$

Notice that when the signal-to-noise ratio increases (decreases), this corresponds to a higher (lower) absolute value of ϑ. In the limit, as $\sigma_\eta^2 \to 0$, then $\vartheta^2 \to 1$, which ensures exact recovery of the state in both cases.[3]

Returning to our endogenous signal extraction problem of (7), we must first find the corresponding innovations associated with observing current and past y_t; thus, we must flip the λ 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, $\mathscr{B}_\lambda(L) \equiv (L-\lambda)/(1-\lambda L)$

$$y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t = (1-\lambda L)\widetilde{Y}(L)e_t \tag{13}$$

$$e_t = \left(\frac{L-\lambda}{1-\lambda L}\right)\varepsilon_t = (L-\lambda)(\varepsilon_t + \lambda\varepsilon_{t-1} + \lambda^2\varepsilon_{t-1} + \cdots) \tag{14}$$

[3]Rondina and Walker (2021) emphasize that while the informational content can be made identical, the dynamics of the two signal extraction problems are drastically different.

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 (13). Note also that *conditional* expectations differ in the e_t and ε_t spaces.

The guess of the equilibrium process (13) must be verified, and uniquely so. This is accomplished by forming the expectation

$$\mathbb{E}[y_{t+1}|\Omega_t^{PI} = \{y_{t-j}\}_{j=0}^{\infty}] = \mathbb{E}[(1-\lambda L)\widetilde{Y}(L)e_{t+1}] = L^{-1}[(1-\lambda L)\widetilde{Y}(L) - \widetilde{Y}_0]e_t \tag{15}$$

which is simply the Wiener-Kolmogorov optimal prediction formula applied to (13). Substituting this expectation into (1) 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 $\beta \in (0,1)$,

$$(z-\lambda)\widetilde{Y}(z) = \frac{zA(z) - \beta A(\beta)\mathscr{B}_\lambda(z)/\mathscr{B}_\lambda(\beta)}{z-\beta} \tag{16}$$

However, there is an additional step that we must take in order to prove that the expectation is consistent with (15) and that the sequence $\{y_{t-j}\}_{j=0}^{\infty}$ does *not* reveal ε_t. We assumed that the endogenous variable is not invertible in λ, this is only true if the RHS of (16) vanishes at $z = \lambda$. This places a 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 ε_t), then the unique rational expectations equilibrium is given by (16). 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 2.1. We have proved the following:

Proposition 1. *Consider the economy described by* (1)–(2) *with expectations given by* $\mathbb{E}[y_{t+1}|\{y_{t-j}\}_{j=0}^{\infty}]$. *If* $\beta \in (0,1)$ *and*

$$A(\lambda) = 0 \tag{17}$$

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

$$y_t = \left(\frac{L(1-\lambda L)\widetilde{A}(L) - \beta(1-\lambda\beta)\widetilde{A}(\beta)}{L-\beta} \right) e_t \tag{18}$$
$$e_t = \left(\frac{L-\lambda}{1-\lambda L} \right) \varepsilon_t$$

If $|\lambda| > 1$*, then the rational expectations equilibrium is unique and given by* (5).

From the perspective of the uninformed agents, the model lives in the e_t space as shown by (18). The model is interpreted as solving the following discounted expectation,

$$y_t = \mathbb{E}_t^U \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t^U \left(\frac{L(1-\lambda L)\widetilde{A}(L)}{L-\beta} \right) e_t \tag{19}$$

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 (18). However, the following corollary derives the equilibrium in the ε_t space.

Corollary 1. *There is an equivalent representation of the equilibrium of Proposition 1 given by*

$$y_t = \left(\frac{L(L-\lambda)\widetilde{A}(L) - \beta(\beta-\lambda)\widetilde{A}(\beta)}{L-\beta} \right) \varepsilon_t - \left[\frac{\beta\widetilde{A}(\beta)(1-\lambda^2)}{1-\lambda L} \right] \varepsilon_t \tag{20}$$

Representation (18) is the equilibrium in e_t space and (20) is the equilibrium in ε_t space. They are equivalent representations of the same equilibrium. Representation (18) 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 (20) 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 (20), note the straightforward decomposition

$$\mathbb{E}_t^U \sum_{j=0}^{\infty} \beta^j x_{t+j} = \mathbb{E}_t^I \left(\sum_{j=0}^{\infty} \beta^j x_{t+j} \right) - \beta\tilde{A}(\beta)(1-\lambda^2) \sum_{k=0}^{\infty} \lambda^k \varepsilon_{t-k} \tag{21}$$

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 $\mathbb{E}_t^I(x_{t+j})$, and a term given by linear combination of past ε_t's that the agents do not observe. Notice that the linear combination is just the dynamic noise term of equation (9) multiplied by the weight $\beta\tilde{A}(\beta)$. As we show below, the representation of Corollary 1 is particularly useful when interpreting equilibrium objects like higher-order beliefs.

2.3 HIERARCHICAL INFORMATION We now introduce heterogeneity in the form of two distinct groups of agents. The first group, in proportion μ, observes the underlying shocks directly, $\Omega_t^{\mu} = \{\varepsilon_{t-j}\}_{j=0}^{\infty}$. 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, $\Omega_t^{1-\mu} = \{y_{t-j}\}_{j=0}^{\infty}$ and are uninformed (U). The corresponding model to be solved is given by

$$y_t = \beta\mu\mathbb{E}^I[y_{t+1}|\Omega^{\mu}] + \beta(1-\mu)\mathbb{E}^U[y_{t+1}|\Omega^{1-\mu}] + x_t \tag{22}$$

$$\Omega_t^{\mu} = \{\varepsilon_{t-j}\}_{j=0}^{\infty}, \qquad \Omega_t^{1-\mu} = \{y_{t-j}\}_{j=0}^{\infty} \tag{23}$$

Following the intuition of the previous section, we assume the endogenous variable is given by the non-invertible process (7), $y_t = (L-\lambda)\widetilde{Y}(L)\varepsilon_t$. In order to prove that such an equilibrium exists, we need to derive a restriction on the exogenous process (x_t) similar to that of (17). Merely assuming that the exogenous process is not invertible following (17) is not sufficient as this restriction only pertains to the representative agent version of Proposition 1. 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 2. *Consider the economy described by* (22) *and* (23). *If* $\beta \in (0,1)$ *and there exists a* $|\lambda| \in (0,1)$ *such that*

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

then the unique rational expectations equilibrium is given by

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

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 2 is similar to that of Proposition 1 with the important difference that now restriction (24) must be satisfied in order for asymmetric information to persist in equilibrium. 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 (22), 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,

$$\begin{aligned} y_t - \beta(1-\mu)\mathbb{E}^U(y_{t+1}) &= \beta\mu\mathbb{E}^I(y_{t+1}) + x_t \\ &= \beta\mu L^{-1}\left[(L-\lambda)Y(L) - \frac{\lambda A(\beta)}{h(\beta)}\right]\varepsilon_t + A(L)\varepsilon_t \end{aligned} \tag{26}$$

where the last equality follows from the proof of Proposition 2 in Appendix A. Equation (26) 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 (26) must be equivalent to e_t in order for equilibrium to be consistent with rational expectations. This will be true if and only if (26) vanishes at $L = \lambda$. Condition (24) ensures that this is the case.

The equilibrium representation of Proposition 2 is algebraically the cleanest because it makes clear $\mu \to 0$ implies convergence to the equilibrium of 1, and as $\mu \to 1$, the equilibrium approaches the fully revealing equilibrium of Section 2.1. However there are equivalent representations which have a more natural economic interpretation, which we state as the following corollaries.

Corollary 2. *The equilibrium described in Proposition 2 has an equivalent representation*

in ε space given by

$$y_t = \left(\frac{LA(L)}{L-\beta}\right)\varepsilon_t - \left(\frac{\beta A(\beta)}{L-\beta}\right)\varepsilon_t - (1-\mu)\beta M^U(\beta)\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_t, \tag{27}$$

where $M^U(\beta) = \frac{A(\beta)}{\beta-\lambda-\mu\beta(1-\lambda^2)}$.

Proof. Follows directly from 2. □

Representation (27) extends Corollary 1 to the heterogeneous agent case. The difference here lies in the third term, where the noise is multiplied by the cumulated term $M^U(\beta)$ rather than $\widetilde{A}(\beta)$. Extending the intuition suggested by Corollary 1, the lag polynomial $M^U(L) = \frac{A(L)}{L-\lambda-\mu\beta(1-\lambda^2)}$ represents the process for the exogenous process x_t *as perceived by the uninformed agents* in the hierarchical information equilibrium. Notice that as μ goes to 0 the perception is once again $\widetilde{A}(L)$. When informed agents are present, $\mu > 0$, the perception changes and as μ increases it gets closer to the actual process $A(L)$ until it exactly coincides for some $\mu = \mu^* < 1$. It is useful to think of the equilibrium perception as being the result of two effects: one due to the simple presence of more informed agents in the market, the other due to the higher-order beliefs that informed agents implicitly form in an equilibrium with heterogeneous information; a point to which we now turn.

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.

Corollary 3. *The equilibrium described in Proposition 2 has an equivalent representation in e space given by*

$$y_t = \frac{1}{L-\beta}\Big\{(1-\lambda L)LH^U(L) - (1-\lambda\beta)\beta H^U(\beta)\Big\}e_t, \tag{28}$$

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 ε space given by

$$y_t = \frac{1}{L-\beta}\Big\{LH^I(L) - \beta H^I(\beta)\Big\}\varepsilon_t \tag{29}$$

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

Proof. Follows directly from Proposition 2. □

Representations (28) and (29) take the more familiar Hansen-Sargent form and show that 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. The modification of the Hansen-Sargent formula is due to the speculative dynamics associated heterogeneous information. By speculative dynamics, we mean that 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.

2.4 DISPERSED INFORMATION The hierarchical informational assumption of the previous section is admittedly extreme. In this section we study equilibria under a more reasonable informational setup. We assume that all agents are identical in terms of the imperfect quality of information they possess. In particular, we assume each agent observes its own particular "window of the world," as in Phelps (1969). Agents observe a noisy signal of the innovation 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.

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 (25), once the parameter that governs the share of informed agents μ 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.

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

$$\varepsilon_{it} = \varepsilon_t + v_{it} \quad \text{with} \quad v_{it} \overset{iid}{\sim} N\left(0, \sigma_v^2\right) \quad \text{for} \quad i \in [0,1] \tag{30}$$

$$\Omega_t^i = \{y_{t-j}, \varepsilon_{i,t-j}\}_{j=0}^{\infty} \quad \text{for} \quad i \in [0,1] \tag{31}$$

The model to be solved is

$$y_t = \beta \int_0^1 \mathbb{E}^i[y_{t+1}|\Omega_t^i]di + x_t \tag{32}$$

Notice that when the noise is driven to zero, $\sigma_v^2 \to 0$, this setup is equivalent to the full information equilibrium of Section 2.1, while an infinite noise, $\sigma_v^2 \to \infty$, yields the incomplete information equilibrium of Section 2.2.

What is unique about this setup is that each agent formulates a forecast by extracting optimally the information from a vector of two signals (y_t, ε_{it}). The basic idea of deriving a fundamental representation developed above 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. We maintain the assumption that the endogenous variable contains exactly one zero inside the unit circle; again, this is without loss of generality. The mapping between innovations and signals is given by

$$\begin{pmatrix} \varepsilon_{it} \\ y_t \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ (L-\lambda)Y(L) & 0 \end{bmatrix} \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix}. \tag{33}$$

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 ε_{it}. The optimal expectation will always be given by the sum of two terms: a linear combination of current and past innovations ε_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

$$\overline{\mathbb{E}}_t(y_{t+1}) = [(L-\lambda)Y(L) + \lambda Y_0]\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2}\varepsilon_t + [(1-\lambda L)Y(L) - Y_0]\frac{\sigma_v^2}{\sigma_\varepsilon^2+\sigma_v^2}e_t \tag{34}$$

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

Proposition 3. *Consider the economy described by* (32) *and* (31). *If* $\beta \in (0,1)$ *and there exists a* $|\lambda| \in [0,1)$ *such that*

$$A(\lambda) - \frac{\tau\beta A(\beta)}{\tau\lambda + (1-\tau)\left(\frac{\beta-\lambda}{1-\lambda\beta}\right)} = 0 \tag{35}$$

then the unique rational expectations equilibrium is given by

$$y_t = \frac{1}{L-\beta}\left\{LA(L) - \beta A(\beta)\left(\frac{\tau\lambda + (1-\tau)\mathscr{B}_\lambda(L)}{\tau\lambda + (1-\tau)\mathscr{B}_\lambda(\beta)}\right)\right\}\varepsilon_t \tag{36}$$

with $\mathscr{B}_\lambda(L) \equiv \frac{L-\lambda}{1-\lambda L}$, $\mathscr{B}_\lambda(\beta) \equiv \frac{\beta-\lambda}{1-\lambda\beta}$ *and* $\tau = \sigma_\varepsilon^2/(\sigma_\nu^2+\sigma_\varepsilon^2)$ *is the signal-to-noise ratio associated with the signal* ε_{it} *in* (31).

Proof. See Appendix A. □

Theorem 1 follows immediately.

Theorem 1. *Let* $\tau \equiv \sigma_\varepsilon^2/(\sigma_\nu^2+\sigma_\varepsilon^2)$ *be the signal-to-noise ratio of (30). The rational expectations equilibrium of Proposition 3 is equivalent to the rational expectations equilibrium of Proposition 2 when* $\mu = \tau$.

The theorem states that in terms of aggregates, the dispersed information setup is identical (i.e., same existence condition (24) and same equilibrium function (25)) to the hierarchical information setup when the signal-to-noise ratio $\tau \equiv \sigma_\varepsilon^2/(\sigma_\nu^2+\sigma_\varepsilon^2)$ is equal to the proportion of informed traders, μ. 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 ε_{it} and a publicly observed signal y_t about the unobserved fundamental ε_t. The optimal behavior—in terms of forecast error minimization—is for the agent to act *as if* the signal ε_{it} contained no noise and thus was equal to the true state ε_t, in measure proportional to the informativeness of the signal τ. At the same time, it is certainly possible that 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 $(1-\tau) \equiv \sigma_\nu^2/(\sigma_\nu^2+\sigma_\varepsilon^2)$. 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 Ω^I with probability τ, and act as if they possess just the public information of the uninformed agents Ω^U with probability $1-\tau$.

While Theorem 1 guarantees equivalence with the hierarchical setup at the aggregate level, there exists important differences between the two equilibria at the individual agent level. 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 4. *The cross-section of beliefs of Theorem 1 is given by*

$$\mathbb{E}_t^i(y_{t+j}) = \mathbb{E}_t^I\left(y_{t+j}\right) - (1-\tau)\,Y_{j-1}\frac{1-\lambda^2}{1-\lambda L}\varepsilon_t - \tau\,Y_{j-1}\frac{1-\lambda^2}{1-\lambda L}v_{it} \text{ for } j = 1,2,.... \tag{37}$$

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

$$\tau^2\left(1-\lambda^2\right)Y_{j-1}^2\sigma_v^2 \text{ for } j = 1,2,.... \tag{38}$$

Proof. See Appendix A. □

If information was complete, 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 ε_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 4 offers an analytical form that can be useful in calibrating key parameters of cross-sectional beliefs.

3 IMPLICATIONS

We now exploit our equivalence results to study higher-order beliefs (HoBs) and information transmission. Theorem 1 permits an interpretation in which agents behave using a mixed strategy approach. We show that this intuition extends to forming higher-order beliefs and derive an analytical characterization. We then employ the various forms of Hansen-Sargent formulas found in Corollaries 1–3 to quantify the effect of endogenous signal extraction on information transmission.

3.1 HIGHER-ORDER BELIEFS

We begin by analyzing higher-order beliefs (HoBs) in the hierarchical equilibrium of Proposition 2. While HoBs in economies with hierarchical information are relatively straightforward, it is the extension to the dispersed-information economy, via Theorem 1, that is our contribution.

Higher-order beliefs in the hierarchical equilibrium follow most naturally from the

equilibrium representations of Corollary 3, which we repeat here for convenience,

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

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 ε space given by

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

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 2, $A(\lambda) - (\mu\beta A(\beta))/(\mu\lambda + (1-\mu)\mathscr{B}_\lambda(\beta)) = 0$ ensures that uninformed agents cannot learn more from the informed forecast error than the space spanned by the e_t process. However, the uninformed do learn from this endogenous signal extraction, which we discuss in more detail below.

In order to derive *higher* order beliefs, we iterate the equilibrium equation forward by one period, $y_{t+1} = \beta\mu\mathbb{E}_{t+1}^I[y_{t+2}] + \beta(1-\mu)\mathbb{E}_{t+1}^U[y_{t+2}] + x_{t+1}$, 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

$$\mu\mathbb{E}_{t+1}^I y_{t+2} + (1-\mu)\mathbb{E}_{t+1}^U y_{t+2} = y_{t+2} + \mu Y_0\lambda\varepsilon_{t+2} - (1-\mu)Y_0\mathscr{B}_\lambda(L)\varepsilon_{t+2} \tag{39}$$

The average market forecast error on the RHS of (39) 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, μ; 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

$$\mathbb{E}_t^I\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \mathbb{E}_t^I y_{t+2} - (1-\mu)Y_0\left(\frac{1-\lambda^2}{1-\lambda L}\right)\lambda\varepsilon_t \tag{40}$$

Hence, the informed agents will always do better (smaller forecast error), if they correct their expectation of the average price according to the forecast errors of the uninformed. Conversely, the uninformed cannot form HoBs 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 $\mathbb{E}_t^U(\overline{\mathbb{E}}_{t+1}y_{t+2}) = \mathbb{E}_t^U y_{t+2}$.

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,

$$\overline{\mathbb{E}}_t\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \overline{\mathbb{E}}_t y_{t+2} - \mu(1-\mu)Y_0\left(\frac{1-\lambda^2}{1-\lambda L}\right)\lambda\varepsilon_t \tag{41}$$

The whole structure of HoBs at any order can be analytically characterized and we direct the interested reader to the general formula in Appendix A. Here we just remark that it is the formation of HoBs that leads directly to the failure of the law of iterated expectations, which is a function of the share of informed agents, μ, and the degree of asymmetric information, as indexed by λ.

It is optimal for informed agents to adjust expectations by correcting the forecast errors of the uninformed; optimal prediction necessitates these adjustments. As we show below, HoBs exist in the dispersed information equilibrium as well. More formally, from Theorem 1, we can write the time-t expectation of agent i of the equilibrium at $t+1$ as

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

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 $\mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}_{t+1}^I y_{t+2} = \mathbb{E}_{it}y_{t+2}$. The law of iterated expectations holding at the single agent level also implies $\mathbb{E}_{it}\mathbb{E}_{it+1}\mathbb{E}_{t+1}^U y_{t+2} = \mathbb{E}_{it}\mathbb{E}_{t+1}^U y_{t+2}$. Therefore

$$\mathbb{E}_{it}\left(\overline{\mathbb{E}}_{t+1}y_{t+2}\right) = \mu\mathbb{E}_{it}y_{t+2} + (1-\mu)\mathbb{E}_{it}y_{t+2} - (1-\mu)Y_0\mathbb{E}_{it}\left(\frac{1-\lambda^2}{1-\lambda L}\right)\varepsilon_{t+1} \tag{42}$$

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

$$(1-\mu)Y_0\mathbb{E}_{it}\left(\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}\right) = Y_0(1-\mu)\mu\lambda\frac{\left(1-\lambda^2\right)}{1-\lambda L}\varepsilon_{it} \tag{43}$$

We have proved the following.

Proposition 5. *Consider the dispersed-information economy described by* (32) *and* (31). *If Proposition 3 holds, then*
i. all agents form higher order beliefs
ii. the average expectations operator does not satisfy the law of iterated expectations.

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 HoBs to be optimal. 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 μ. 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, (42) 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 2 and 3, specifically Equation (35), to determine the required τ or μ necessary to completely reveal the underlying shock sequence, ε^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 3 in the (β,θ) 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 2.) The figure is built around the following corollary to Proposition 3.

Corollary 4. *Consider the dispersed-information economy described by* (32) *and* (31) *of Proposition 3 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* (β, θ) *space according to the following restrictions:*

(4.a) If $\theta \leq 1$, *a dispersed information equilibrium does not exist and the model is characterized by the full-information counterpart of Section 2.1.*

(4.b) If $\theta > 1$, *a dispersed information equilibrium exists for any* $\tau > 0$ *and* $\rho \geq 0$ *if*

$$\theta \geq \left(\frac{1}{1-\beta(1+\rho)}\right) \tag{4.b}$$

(4.c) If $\theta > 1$ *and (4.b) is not satisfied, a dispersed information equilibrium exists for signal-to-noise ratio* τ *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. □

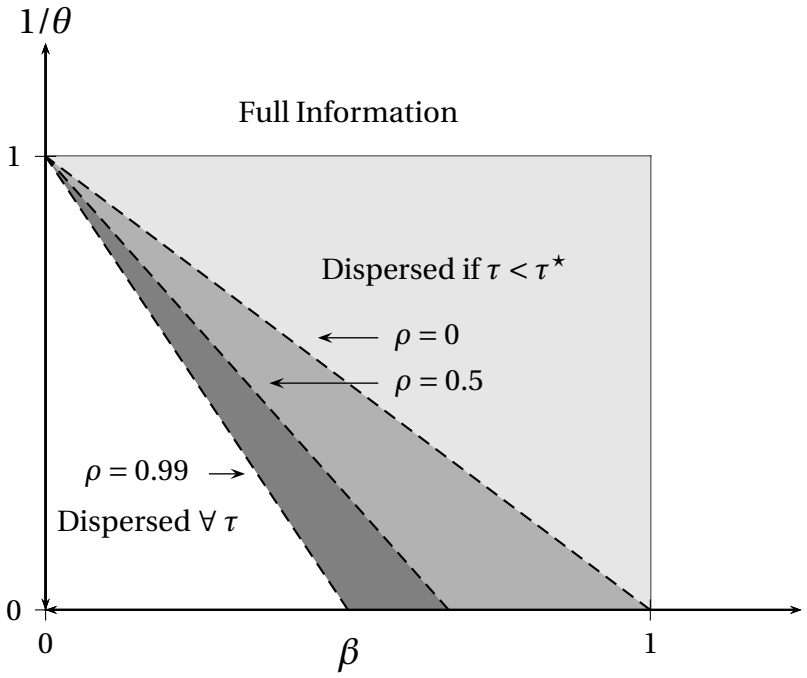

Figure 1: (β, θ) Existence Space. Existence of Dispersed and Full-Information Equilibria following Proposition 3 for $x_t = \rho x_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$. Equilibria to the right of the dashed line preserve heterogeneity in information if and only if $\tau < \tau^\star$.

Three points are noteworthy. First, as is evident from the figure and condition (4.a), if $\theta \leq 1$, the endogenous variable fully reveals the underlying shock, ε_t, and the equilibrium is consistent with the complete information case of Section 2.1. With $\theta \leq 1$, confounding dynamics are not present in the exogenous process and x_t is fundamental for ε_t. Second, from condition (4.c) 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 2.2 demonstrate that an increase in θ may be interpreted as an increase in the noise associated with the endogenous signal extraction problem. The information content of the endogenous variable is sufficiently small that no matter how informative the exogenous signal, the full information equilibrium cannot be learned. How the discount factor β alters the space of existence is similar to that of the serial correlation parameter ρ, which is the final point to be made. As the serial correlation in the x_t process increases and β increases, it is more difficult to preserve dispersed information, *ceteris paribus* (the dashed line shifts to the left as ρ increases from 0 to 0.99). An increase in β and ρ leads to a longer lasting effect of current information. This results in a higher $|\lambda|$ and a decrease in the informational discrepancy between fully informed and uninformed agent types.

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 2 and 3 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 4. 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

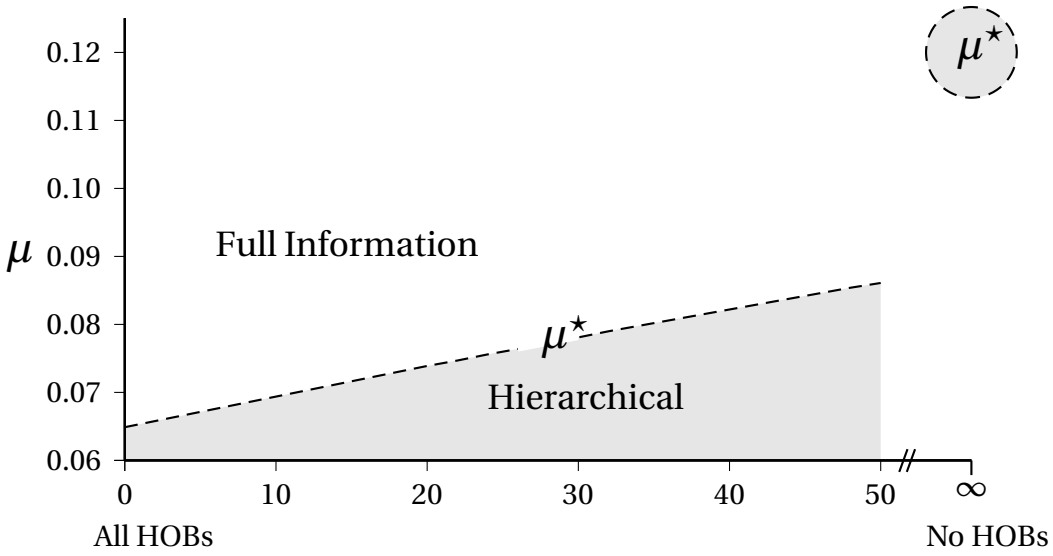

Figure 2: Existence space for the hierarchical information equilibrium as higher-order beliefs are removed from the expectation of informed agents: $x_t = 0.8x_{t-1} + \varepsilon_t + \sqrt{11}\varepsilon_{t-1}$, $\beta = 0.985$.

is declining as fewer HoBs are being formulated. When we impose the law of iterated expectations on the entire dynamic structure (No HoBs or ∞ on the x-axis for Figure 2), the number of informed agents or the informativeness of the exogenous signal can nearly double (from 0.065 to 0.122) without fully revealing all underlying shocks.

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.

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.

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.

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

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.

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

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., "Interpreting Economic Time Series," *Journal of Political Economy*, 1981, *89* (2), 213–248.

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

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.

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

A PROOFS

A.1 PROOF OF EQUATIONS (11)–(12)

We need to show that the representations (8) and (10) are equivalent in terms of unconditional forecast error variance

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

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 ε, $\mathbb{E}(\varepsilon_t|\mathscr{S}^t) = \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_\eta^2}\right)\mathscr{S}_t$ and therefore,

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|\mathscr{S}^t\right)\right)^2\right] = \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_\varepsilon^2 + \sigma_\eta^2} \tag{45}$$

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

$$s_t = (L-\vartheta)\left(\frac{1-\vartheta L}{L-\vartheta}\right)\left(\frac{L-\vartheta}{1-\vartheta L}\right)\varepsilon_t = (1-\vartheta L)e_t \tag{46}$$

$$e_t = \left(\frac{L-\vartheta}{1-\vartheta L}\right)\varepsilon_t \tag{47}$$

Given that (46) 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

$$\mathbb{E}(\varepsilon_t|e^t) = \mathbb{E}(\varepsilon_t|s^t) \tag{48}$$

To show (48) notice that (47) can be written as

$$\varepsilon_t = C(L)e_t = \left[\frac{1-\vartheta L}{L-\vartheta}\right]e_t = \left[\frac{L^{-1}-\vartheta}{1-\vartheta L^{-1}}\right]e_t = (L^{-1}-\vartheta)\sum_{j=0}^{\infty}\vartheta^j e_{t+j} \tag{49}$$

Thus, while (46) 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,

$$\mathbb{E}(\varepsilon_t|e^t) = \left[C(L)\right]_+ e_t = -\vartheta e_t = \left(\frac{-\vartheta}{1-\vartheta L}\right)s_t = \mathbb{E}(\varepsilon_t|s^t) \tag{50}$$

We must now calculate

$$\mathbb{E}\left[\left(\varepsilon_t - \mathbb{E}\left(\varepsilon_t|s^t\right)\right)^2\right] = \mathbb{E}\left(\varepsilon_t^2\right) + \mathbb{E}\left(\varepsilon_t|s^t\right)^2 - 2\mathbb{E}\left(\varepsilon_t \mathbb{E}\left(\varepsilon_t|s^t\right)\right) \tag{51}$$

$$= \sigma_\varepsilon^2 + \vartheta^2\sigma_\varepsilon^2 - 2\mathbb{E}(\varepsilon_t\left(\varepsilon_t|s^t\right)) \tag{52}$$

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,

$$\mathbb{E}(\varepsilon_t e_t) = \frac{-\vartheta\sigma_\varepsilon^2}{2\pi i}\oint \frac{z-\vartheta}{1-\vartheta z}\frac{dz}{z} = \sigma_\varepsilon^2\left[\lim_{z\to 0}\frac{z-\vartheta}{1-\vartheta z}\right] = \vartheta^2\sigma_\varepsilon^2 \tag{53}$$

Equations (52) and (53) give the desired result

$$\mathbb{E}\left[\left(\varepsilon_t - E\left(\varepsilon_t|x^t\right)\right)^2\right] = \left(1-\vartheta^2\right)\sigma_\varepsilon^2 \tag{54}$$

Equating (54) and (45) concludes the proof,

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

A.2 PROOF OF PROPOSITION 2 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 ε_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

$$\begin{aligned}(z-\beta)Y(z) &= \frac{1}{z-\lambda}\left\{zA(z)+\beta Y_0[\mu\lambda-(1-\mu)\mathscr{B}_\lambda(z)]\right\}\\ &= \frac{1}{z-\lambda}\left\{zA(z)+\beta Y_0 h(z)\right\}\end{aligned} \tag{55}$$

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 (24). Substituting this into (55) delivers (25).

A.3 PROOF OF PROPOSITION 3 Similar to solving the previous model, the first step in the proof of Proposition 3 is to obtain an innovations representation for the signal vector (ε_{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 ε_t and the noise v_{it} is

$$\begin{pmatrix}\varepsilon_{it}\\ y_t\end{pmatrix}=\begin{pmatrix}\sigma_\varepsilon & \sigma_v\\ (L-\lambda)Y(L) & 0\end{pmatrix}\begin{pmatrix}\hat{\varepsilon}_t\\ \hat{v}_{it}\end{pmatrix}=\Gamma(L)\begin{pmatrix}\hat{\varepsilon}_t\\ \hat{v}_{it}\end{pmatrix} \tag{56}$$

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

$$\begin{pmatrix}\varepsilon_{it}\\ y_t\end{pmatrix}=\Gamma^*(L)\begin{pmatrix}w_{it}^1\\ w_{it}^2\end{pmatrix} \tag{57}$$

As with the hierarchical case, we must use Blaschke factors to flip the non-fundamental root, λ, to outside the unit circle. However, we must also employ a Gram-Schmidit type orthogonalization (W_λ) 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 (57)

$$\mathbb{E}(y_{t+1}|\varepsilon_i^t, y^t) = \left[\Gamma_{21}^*(L) - \Gamma_{21}^*(0)\right] L^{-1} w_{it}^1 + \left[\Gamma_{22}^*(L) - \Gamma_{22}^*(0)\right] L^{-1} w_{it}^2. \tag{58}$$

It is straightforward to show that

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

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

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

The average market expectation is then

$$\bar{\mathbb{E}}(y_{t+1}) = \left[(L-\lambda)Y(L) + \lambda Y_0\right] L^{-1} \tfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2} \hat{\varepsilon}_t + \left[(1-\lambda L)Y(L) - Y_0\right] L^{-1} \tfrac{\sigma_v^2}{\sigma_\varepsilon^2+\sigma_v^2} \frac{L-\lambda}{1-\lambda L} \hat{\varepsilon}_t \tag{59}$$

Now, if we let

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

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

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

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

$$(z-\lambda)Q(z) = \beta\tau z^{-1}[(z-\lambda)Q(z)+\lambda Q_0] + \beta(1-\tau)z^{-1}\left[(z-\lambda)Q(z)+Q_0\frac{\lambda-L}{1-\lambda L}\right] + A(z) \tag{60}$$

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

$$\begin{aligned}(z-\beta)Q(z) &= \frac{1}{z-\lambda}\left[zA(z) + \beta Q_0\left(\tau\lambda + (1-\tau)\frac{\lambda-z}{1-\lambda z}\right)\right] \\ &= \frac{1}{z-\lambda}\left[zA(z) + \beta Q_0 h(z)\right]\end{aligned} \tag{61}$$

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 λ can be then written as

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

which is (35). Substituting this into (61) delivers (36), which completes the proof.

A.4 PROOF OF PROPOSITION 4 Once the analytic form for $\Gamma^*_{21}(L)$ and $\Gamma^*_{22}(L)$ are known from Proposition 3, 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 (58) and collecting the terms that constitute (59), one gets

$$\begin{aligned}\mathbb{E}(y_{t+1}|\varepsilon_i^t, y^t) &= \bar{\mathbb{E}}(y_{t+1}) + \frac{\sigma_\varepsilon}{\sigma_\varepsilon^2+\sigma_v^2}L^{-1}[(L-\lambda)Y(L) + \lambda Y_0 - (L-\lambda)Y(L) + Y_0\frac{L-\lambda}{1-\lambda L}]\sigma_v\hat{v}_{it} \\ &= \bar{\mathbb{E}}(y_{t+1}) + \frac{\sigma_\varepsilon}{\sigma_\varepsilon^2+\sigma_v^2}L^{-1}[\lambda Y_0 + Y_0\frac{L-\lambda}{1-\lambda L}]\sigma_v\hat{v}_{it} \\ &= \bar{\mathbb{E}}(y_{t+1}) + \mu Y_0\frac{1-\lambda^2}{1-\lambda L}v_{it},\end{aligned} \tag{62}$$

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 σ_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 2. For $j = 1$, the time $t+1$ average expectation at $t+2$ is given by

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

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

$$\mathbb{E}^I_t \overline{\mathbb{E}}_{t+1} y_{t+2} = \mathbb{E}^I_t y_{t+2} + \mu\lambda Y_0 \mathbb{E}^I_t \varepsilon_{t+2} - Y_0(1-\mu)\mathbb{E}^I_t \mathscr{B}_\lambda(L)\varepsilon_{t+2}. \tag{64}$$

Clearly $\mathbb{E}^I_t \varepsilon_{t+2} = 0$, whereas the expectation in the last term of (64) is given by

$$\mathbb{E}^I_t \mathscr{B}_\lambda(L)\varepsilon_{t+2} = L^{-2}\{\mathscr{B}_\lambda(L) - \mathscr{B}_\lambda(0) - \mathscr{B}_\lambda(1)L\}\varepsilon_t \tag{65}$$

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

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

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

$$\mathbb{E}^I_t \overline{\mathbb{E}}_{t+1} y_{t+2} = \mathbb{E}^I_t y_{t+2} - (1-\mu)Y_0 \lambda \Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_t \tag{66}$$

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, $\mathbb{E}^U_t \varepsilon_{t+2} = 0$; however, the second term now is $\mathbb{E}^U_t \mathscr{B}_\lambda(L)\varepsilon_{t+2} = 0$ because, by definition, $\mathscr{B}_\lambda(L)\varepsilon_{t+2}$ is not in the information set of the uninformed agents at time t. Hence $\mathbb{E}^U_t \overline{\mathbb{E}}_{t+1} y_{t+2} = \mathbb{E}^U_t y_{t+2}$: the uninformed are *not* forming higher-order expectations.

Applying the above results to the market forecast of the market forecast one gets

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}y_{t+2} = \mu\mathbb{E}_t^I\overline{\mathbb{E}}_{t+1}y_{t+2} + (1-\mu)\mathbb{E}_t^U\overline{\mathbb{E}}_{t+1}y_{t+2} \quad = \overline{\mathbb{E}}_t y_{t+2} - \mu(1-\mu)Y_0\lambda\Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_t, \quad (67)$$

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 (63),

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

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

$$\begin{aligned}\mathbb{E}_{t+1}^U y_{t+3} &= \Big[\frac{(1-\lambda L)Y(L)}{L^2}\Big] + \mathscr{B}_\lambda(L)\varepsilon_{t+1} \\ &= L^{-2}[(L-\lambda)Y(L) - (Y_0 + (Y_1 - \lambda Y_0)L)\mathscr{B}_\lambda(L)]\varepsilon_{t+1} \end{aligned} \quad (68)$$

For the informed

$$\begin{aligned}\mathbb{E}_{t+1}^I\overline{\mathbb{E}}_{t+2}y_{t+3} &= \mathbb{E}_{t+1}^I y_{t+3} + \mu Y_0\lambda\mathbb{E}_{t+1}^I\varepsilon_{t+3} - (1-\mu)Y_0\mathbb{E}_{t+1}^I\mathscr{B}_\lambda(L)\varepsilon_{t+3} \\ &= L^{-2}[(L-\lambda)Y(L) + \lambda Y_0 - (Y_0 - \lambda Y_1)L]\varepsilon_{t+1} - (1-\mu)Y_0\lambda\Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_{t+1}. \end{aligned} \quad (69)$$

Combining (68) and (69) gives

$$\begin{aligned}\overline{\mathbb{E}}_{t+1}\overline{\mathbb{E}}_{t+2}y_{t+3} = y_{t+3} &+ \mu\{\lambda Y_0 - (Y_0 - \lambda Y_1)L\}\varepsilon_{t+3} - \mu(1-\mu)Y_0\lambda\Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_{t+1} \\ &-(1-\mu)[Y_0 + (Y_1 - \lambda Y_0)L]\mathscr{B}_\lambda(L)\varepsilon_{t+3} \end{aligned} \quad (70)$$

Following the same argument that we used for the first order expectations it is easy to conclude that the uninformed's expectations of (70) are just

$$\mathbb{E}_t^U\overline{\mathbb{E}}_{t+1}\overline{\mathbb{E}}_{t+2}y_{t+3} = \mathbb{E}_t^U y_{t+3} \quad (71)$$

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 re-

ceive 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

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

Therefore the average expectation is

$$\overline{\mathbb{E}}_t\overline{\mathbb{E}}_{t+1}\overline{\mathbb{E}}_{t+2}y_{t+3} = \overline{\mathbb{E}}_t y_{t+3} - (1-\mu)(Y_0\mu^2\lambda^2 + Y_1\mu\lambda)\Big(\frac{1-\lambda^2}{1-\lambda L}\Big)\varepsilon_t. \tag{72}$$

Comparing this to (67) 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

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

A.6 PROOF OF PROPOSITION 5

We begin by noticing that

$$\mathbb{E}_{it}\overline{\mathbb{E}}_{t+1}y_{t+2} = \mu\mathbb{E}_{it}\mathbb{E}_{t+1}^I y_{t+2} + (1-\mu)\mathbb{E}_{it}\mathbb{E}_{t+1}^U y_{t+2}. \tag{73}$$

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

$$\mathbb{E}_{it}\overline{\mathbb{E}}_{t+1}y_{t+2} = \mu\mathbb{E}_{it}y_{t+2} + (1-\mu)\mathbb{E}_{it}y_{t+2} - (1-\mu)Y_0\mathbb{E}_{it}\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}. \tag{74}$$

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

$$\mathbb{E}\left(\frac{1-\lambda^2}{1-\lambda L}\varepsilon_{t+1}|\varepsilon_i^t, y^t\right) = \mathbb{E}\left(\frac{L-\lambda}{1-\lambda L}\varepsilon_{t+2}|\varepsilon_i^t, y^t\right). \tag{75}$$

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

$$\mathbb{E}\left(\frac{L-\lambda}{1-\lambda L}\varepsilon_{t+2}|\varepsilon_i^t, y^t\right) = \mu\lambda\frac{\left(1-\lambda^2\right)}{1-\lambda L}\varepsilon_{it}. \tag{76}$$

where $\mu \equiv \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_v^2}$. Remember that we defined

$$e_t = \mathscr{B}(L)\varepsilon_t. \tag{77}$$

From Theorem 1 in Rondina (2009) we know that

$$\left[\begin{array}{cc} \pi_1(L) & \pi_2(L) \end{array}\right] = \left[L^{-2}g_{e,(\varepsilon,y)}(L)\left(\Gamma^*(L^{-1})^T\right)^{-1}\right]_+ \Gamma^*(L)^{-1} \tag{78}$$

where $\Gamma^*(L)$ and $\left(w_{it}^1, w_{it}^2\right)$ are defined in (57) 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

$$g_{e,(\varepsilon,y)}(L) = \left[\begin{array}{cc} \mathscr{B}(L)\sigma_\varepsilon^2 & \mathscr{B}(L)\left(L^{-1}-\lambda\right)Y\left(L^{-1}\right)\sigma_\varepsilon \end{array}\right].$$

Plugging in the explicit forms and solving out the algebra

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

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 $p(L)$ 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

$$\left[L^{-2}\frac{L-\lambda}{1-\lambda L}\right]_+ = \frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}, \tag{79}$$

Summarizing we have shown that

$$\left[\begin{array}{cc} \pi_1(L) & \pi_2(L) \end{array}\right] = \frac{1}{\sqrt{\sigma_\varepsilon^2+\sigma_v^2}}\left[\begin{array}{cc} \frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}\sigma_\varepsilon^2 & 0 \end{array}\right]\Gamma^*(L)^{-1}.$$

Notice that

$$\Gamma^*(L)^{-1}\begin{bmatrix} \varepsilon_{it} \\ y_t \end{bmatrix} = \begin{bmatrix} w^1_{it} \\ w^2_{it} \end{bmatrix}$$

so that

$$\mathbb{E}\left(y_{t+2}|\varepsilon^t_i, y^t\right) = \begin{bmatrix} \pi_1(L) & \pi_2(L) \end{bmatrix} \begin{bmatrix} \varepsilon_{it} \\ y_t \end{bmatrix} = \frac{1}{\sqrt{\sigma^2_\varepsilon+\sigma^2_v}} \frac{\lambda\left(1-\lambda^2\right)}{1-\lambda L}\sigma^2_\varepsilon w^1_{it}.$$

From the proof of Theorem 3 we know that $w^1_{it} = \frac{1}{\sqrt{\sigma^2_\varepsilon+\sigma^2_v}}(\varepsilon_t + v_{it})$, which, once substituted in the above expression, completes the proof of statement (i). The proof can be generalized to expectations of order higher than 1. For statement (ii) the proof follows exactly the proof of Proposition 3 since it concerns only aggregate variables, which we know from the proof of Theorem 1 follow the same patter as those of the hierarchical case.

A.7 PROOF OF COROLLARY 4 The proof follows immediately from the restriction (24). Condition (4.a) is derived by taking the limit of (24) 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 (24) and solving for μ gives condition (4.c). When $\lambda = -1$, the equilibrium converges to the full-information case. Setting $\mu^\star$ equal to unity and solving for θ gives condition (4.b).