Rational Expectations Models with Higher-Order Beliefs*

Zhen Huo
Yale University

Naoki Takayama
Hitotsubashi University

May 5, 2023

Abstract

We develop a method of solving rational expectations models with dispersed information and dynamic strategic complementarities. In these types of models, the equilibrium outcome hinges on an infinite number of higher-order expectations which require an increasing number of state variables to keep track of. Despite this complication, we prove that the equilibrium outcome always admits a finite-state representation when the signals follow finite ARMA processes. We also show that such finite-state result may not hold with endogenous information aggregation. We further illustrate how to use the method to derive comparative statics, characterize equilibrium outcomes in HANK-type network games, reconcile with empirical evidence on expectations, and integrate incomplete information with bounded rationality in general equilibrium.

Keywords: Higher-order expectations, dynamic complementarity, general equilibrium.

*Huo: zhen.huo@yale.edu. Takayama: ntakayama@ier.hit-u.ac.jp. Acknowledgements: We are thankful for our advisers José-Víctor Ríos-Rull, Ellen McGrattan, Jonathan Heathcote, and Elena Pastorino for inspiration, encouragement, and support. We are grateful for comments from George-Marios Angeletos, Anmol Bhandari, V.V. Chari, Kyle Jurado, Patrick Kehoe, Kris Nimark, Fabrizio Perri, Giacomo Rondina, Thomas Sargent, Todd Walker, Pierre-Olivier Weill, Kei-Mu Yi, and Venky Venkateswaran. We also thank for the comments from participants at various workshops and conferences. Zhen Huo thanks the University of Minnesota Doctoral Dissertation Fellowship for financial support. Naoki Takayama thanks the JSPS KAKENHI 21K13256 for financial support.

1. Introduction

In many economic problems, agents' decisions hinge on expectations about other agents' decisions. For example, when a manager resets their product's price, she needs to take into account their competitors' pricing strategy in the future; when a consumer plans how much to consume, she needs to form expectations about her future income, which in turn depends on future aggregate expenditure. With heterogeneous information among agents, the coordination motive makes aggregate outcomes depend on higher-order uncertainty, or beliefs about others' beliefs. In these environments, what are the macroeconomic effects of incomplete information? How do they interact with the general equilibrium consideration?

In this paper, we propose a method that helps solve and characterize the equilibrium in models featuring dispersed information and dynamic strategic interactions among agents.[1] We prove that when signals follow ARMA processes, aggregate outcomes permit a tractable finite-state representation despite the vast complexity of higher-order expectations. We show that the interaction between informational frictions and dynamic complementarities can be summarized a single equation, the roots of which shape the propagation mechanism in the equilibrium. These results further yield a sequence of applied lessons.

Framework. We consider the following baseline framework, in which the individual agent i's best response is

$$a_{it} = \mathbb{E}_{it}[\xi_t] + \mathbb{E}_{it}[\beta(L)a_{it}] + \mathbb{E}_{it}[\gamma(L)a_t],$$

where ξ_t is some exogenous economic fundamental, a_t is the aggregate outcome. The lag operator function $\beta(L)$ captures the dependence on the agent's own past and future actions. Importantly, we allow $\gamma(L)$ to capture dependence on others actions' or the dynamic complementarities, and agents do not share a common information set. This framework nests a variety of applications with incomplete information, including the monetary model as in Maćkowiak and Wiederholt (2009), the asset pricing model as in Allen, Morris, and Shin (2006), the New Keynesian model as in Nimark (2008), and so on. In a macroeconomic setting, $\beta(L)$ effectively summarizes the partial equilibrium (PE) consideration, while $\gamma(L)$ summarizes the general equilibrium (GE) consideration.

The joint presence of incomplete information and strategic interactions implies that the outcome relies not only on first-order expectations, but also on higher-order expectations. To accommodate higher-order uncertainty is important in macroeconomics: it makes room for forces akin to animal spirits (Lorenzoni, 2009), modifies relative strength between GE and PE (Angeletos and Lian, 2018), induces dynamics that are empirically relevant (Woodford, 2003), and many others.

However, when strategic interactions are intertemporal or heterogeneous cross-sectionally, the types of higher-order expectations involved can be quite complex. With persistent information, to keep track of these higher-order expectations may require the entire history of signals, known as the

[1] Throughout the paper, we focus on linear models with Gaussian shocks.

infinite regress problem (Townsend, 1983). Whether there exists a small set of sufficient statistics to summarize the relevant information in equilibrium is unknown ex ante, which is in contrast with models with perfect information where the state variables are typically straightforward to identify. This is why these types of models are challenging to solve and characterize.

Inference Problem. To overcome this difficulty, we propose a joint use of the Wiener-Hopf prediction formula and the Kalman filter. The Wiener-Hopf prediction formula allows one to forecast the aggregate outcome without specifying the set of state variables (Sargent and Hansen, 1981; Whiteman, 1983; Kasa, 2000; Rondina, 2008; Kasa, Walker, and Whiteman, 2014). However, it still requires a fundamental representation of the signal process so that a new set of transformed shocks contain the same amount of information as the signals.[2] Under the assumption that signals follow ARMA processes, the steady-state Kalman filter is ready to provide such representation. Our contribution is bridging the Kalman filter with the Wiener-Hopf prediction formula, which makes the inference problem tractable even for complicated signal structures and allows us to derive the theoretical results in a general model environment.

With the help of these tools, we first derive a compact formula for the forecast of an arbitrary random variable, which overcomes the difficulty in implementing the annihilation operator for rational functions. In the end, solving for the equilibrium policy rule boils down to finding a fixed point in the space of analytic functions. The transformation from the time domain to the frequency domain avoids the task of looking for an infinite sequence of coefficients that map the entire history of signals to the current outcome, and it yields a much simpler characterization of the equilibrium than one may expect.

Finite-State Representation. Our main result is that the equilibrium outcome always permits a finite-state representation, despite all the complications due to higher-order expectations. This result holds when signals follow ARMA processes, a condition that can be guaranteed when the information flow is *exogenously* determined. The key observation is that recording all higher-order expectations indeed requires an infinite number of state variables, but the equilibrium outcome only depends on a weighted average of them which surprisingly collapses to a low dimensional object. We provide an explicit solution formula, and the associated condition that determines the uniqueness and existence of the equilibrium.

An important feature of models with incomplete information is the sluggish response due to the fact that higher-order expectations are more anchored than first-order expectations (Woodford, 2003; Nimark, 2008; Angeletos and Huo, 2021). We show that the parameters that characterize the additional persistence of the outcome are the outside roots of a single polynomial equation involving four elements: (1) the informational friction captured by the covariance structure of the signal process, (2) the information incompleteness captured by the presence of private signals, (3) the partial equilibrium

[2]The original representation of the signal process typically has the feature that shocks contain more information than signals, so agents need to solve a signal extraction problem to infer the underlying shocks.

consideration $\beta(L)$, and (4) the general equilibrium consideration $\gamma(L)$. These roots hinge on the interaction between the informational friction and the coordination motive only if elements (2) and (4) are simultaneously present, or when higher-order expectations play a role in determining the outcome. The properties of the forecasts determined by informational friction itself do not directly transmit to the endogenous outcome (which would be the case when information heterogeneity or coordination motive was eliminated). Instead, the underlying informational friction may loom larger according to the dynamic complementarities and the amount of information in the public domain relative to that in the private domain.

One may note that the nature of the competitive equilibrium concept allows agents to treat the law of motion of the aggregate outcome as given, without considering all the higher-order expectations. Our strategy in solving the equilibrium follows this logic by solving for the fixed point directly. Ex post, the finite-state result implies a purely statistical observation: a weighted sum of infinite higher-order expectations obeys a finite-order ARMA process. This result reconciles the disparity between the complex higher-order expectations considered by economists and the simple equilibrium law of motion considered by agents within the model. With static strategic concern, this type of dimension reduction takes a particularly sharp form where the sum of higher-order expectations is identical to a more noisy first-order expectation (Huo and Pedroni, 2020).[3]

Our main result can also be extended to a class of models with network structure. In this context, agents may differ from each other in terms of how their payoff depends on the activities of different groups as well as their signal structures. This connects to the literature on network games with incomplete information but static information (Bergemann, Heumann, and Morris, 2017; Golub and Morris, 2019), and a growing literature that emphasizes the interaction between the network structure of the macroeconomy and the informational friction (Auclert, Rognlie, and Straub, 2020; La'O and Tahbaz-Salehi, 2020; Chahrour, Nimark, and Pitschner, 2021).

Applications. Through a sequence of applications, we demonstrate how our results can help develop applied lessons and illustrate the role of higher-order uncertainty in shaping equilibrium outcomes. Our first application shows that our method facilitates closed-form solutions and proofs of comparative statistics. We revisit the classical beauty-contest models à la Woodford (2003) and Angeletos and La'O (2010) with both private and public signals. In response to both the fundamental shock and the common noise, the outcome displays additional persistence, which encapsulates the effects of all the higher-order expectations. Thanks to the analytical solution, we prove that this endogenous persistence is increasing in both the informational friction and the degree of strategic complementarity. We further prove that the volatility of the outcome driven by the fundamental is decreasing in the degree of complementarity, but the outcome driven by the common noise is instead increasing in the degree of complementarity.

[3]With static complementarity, the infinite sum of higher-order expectations is equivalent to the first-order expectation with the precision of private shocks discounted by the degree of complementarity.

In the second application, we extend our analysis to a HANK-type model with incomplete information (Auclert, Rognlie, and Straub, 2020; Angeletos and Huo, 2021). In this model, the interdependence among different consumers amounts to a network game. The analytical solution allows us to characterize the interaction of three types of heterogeneity among consumers: marginal propensity to consume (MPC), income exposure to output, and informational frictions about the underlying fundamental. We show sequentially that (1) with common information structure among consumers, increasing high MPC consumers' income exposure to output amplifies the effects of incomplete information; (2) fixing the average informational friction in the economy, a reduction of the information received by high MPC consumers has a larger quantitative bite. Notably, these results can be understood via a small number of statistics that govern both the impact and dynamic effects, which complement the recent work that studies the full-blown HANK model without imposing full-information and rational expectations (Pfäuti and Seyrich, 2022; Guerreiro, 2022; Gallegos, 2023).

In the third application, we compare our solution under rational expectations with that under certain bounded rationality, and discuss how to distinguish them using the evidence on forecasts. We extend the decentralized-trading model as in Angeletos and La'O (2013) to an environment with persistent sentiment shocks. With rational expectations, the persistence of the equilibrium outcome is always smaller than that of the exogenous sentiment, a footprint of the general equilibrium force. In Angeletos and La'O (2013), a heterogeneous prior approach is adopted,[4] which significantly simplifies the dynamics of the higher-order expectations at the cost of individual rationality. In this case, the persistence of the equilibrium outcome is always identical to that of the sentiment shock. To distinguish these alternative approaches, we consider the regressions proposed by Coibion and Gorodnichenko (2015) and Bordalo, Gennaioli, Ma, and Shleifer (2020) that estimate the predictability of forecast error using forecast revision at the aggregate and individual level, respectively. With rational expectations, the regression coefficient at the aggregate level is positive and is larger than that at the individual level, broadly consistent with the empirical regularities. However, under heterogeneous prior, these two regression coefficients become identical and negative.

In the forth application, we show how our method can help integrate incomplete information with belief distortions in a general equilibrium setting. To illustrate, we extend the beauty-contest model to allow diagnostic expectation à la Bordalo, Gennaioli, Ma, and Shleifer (2020). In Bordalo, Gennaioli, Ma, and Shleifer (2020), the process of the variable to be forecast is specified exogenously. In contrast, in our environment, the outcome is determined in equilibrium and is the result of agents' expectations. Despite the deviation from strict rationality, the finite-state result still applies, and the interaction between the general equilibrium consideration and the belief distortion jointly determines the equilibrium outcome. Relative to the case with rational expectations, the diagnostic expectation modifies both the impact response and the propagation later on, but nevertheless leaves the long-term persistence the same. This approach complements the literature that studies the general equilib-

[4]With heterogeneous prior, shocks are perfectly observed but agents believe that all others observe biased signals.

rium implication of diagnostic expectations but focuses on environments with common information (L'Huillier, Singh, and Yoo, 2021; Bordalo, Gennaioli, Shleifer, and Terry, 2021; Bianchi, Ilut, and Saijo, 2021).

Endogenous Information. Lastly, we consider the environment with endogenous information, that is, the signals may contain aggregate outcomes determined in equilibrium, such as outputs or prices. Our preferred interpretation of the endogenous information equilibrium consists two parts: (1) competitive agents take the signal process as exogenously given, and choose their actions according to their best response function; (2) the law of motion of the implied aggregate outcome is the same as the perceived one that enters the signal process. The first part corresponds to an exogenous-information equilibrium and therefore our previous results can still apply, while the second part imposes an additional fixed point problem. Conceptually, the endogenous-information is a particular exogenous-information equilibrium, where the signal process satisfies additional restrictions. However, with such endogenous signal processes, the finite-state result may not be true. We offer an example which naturally extends previous models, but the equilbliurm law of motion cannot be represented by a finite ARMA process.[5]

Related literature. This paper is most closely related to the literature that studies dynamic linear models with the frequency domain approach. With complete information, Sargent and Hansen (1981) and Whiteman (1983) provide a general solution formula for linear models with rational expectations. With incomplete information, a number of papers (Kasa, 2000; Acharya, 2013; Kasa, Walker, and Whiteman, 2014; Tan and Walker, 2015; Rondina and Walker, 2021; Acharya, Benhabib, and Huo, 2021) explore models with endogenous information aggregation. This line of work utilizes the properties that the number of signals is the same as the number of shocks, and obtain the fundamental representation by flipping the inside root using Blascke matrices. Taub (1989), Rondina (2008) and Miao, Wu, and Young (2021) instead consider the case where there are more shocks than signals, and obtain the fundamental representation through a canonical factorization (Rozanov, 1967). Our approach instead relies on the steady-state Kalman filter to obtain the fundamental representation, which is essential for the derivation of our theoretical results in a general setting. Furthermore, it has an attractive feature that even for large scale state-space systems, the Kalman filter problem can be solved in a fast and robust way.

This paper is also closely related to a large literature at the interaction between macroeconomics and incomplete information, which could be traced back to Lucas (1972). On the theoretical side, it builds on the studies on beauty-contest games (Morris and Shin, 2002; Angeletos and Pavan, 2007; Bergemann and Morris, 2013) and extends the analysis to models with intertemporal coordination and persistent information. On the applied side, it complements a large amount of work on the effects of monetary shocks (Woodford, 2003; Mankiw and Reis, 2002; Maćkowiak and Wiederholt, 2009;

[5]In a different envrionment, Makarov and Rytchkov (2012) also show that the equilibrium does not admit a Markovian dyanmics.

Hellwig and Venkateswaran, 2009; Melosi, 2016), the non-fundamental driven aggregate fluctuations (Lorenzoni, 2009; Angeletos and La'O, 2010; Barsky and Sims, 2012; Angeletos and La'O, 2013; Nimark, 2014; Benhabib, Wang, and Wen, 2015; Huo and Takayama, 2022; Chahrour and Jurado, 2018), and the propagation mechanism of business cycle fluctuations with imperfect information (Bacchetta and Wincoop, 2006; Graham and Wright, 2010; Venkateswaran, 2014; Maćkowiak and Wiederholt, 2015; Angeletos and Lian, 2018; Chahrour and Gaballo, 2016).[6] The equilibrium characterization provided in this paper can also be applied to many of the model economies in the aforementioned papers. Particularly, the finite-state result implies that the guess-and-verify approach used in Woodford (2003) and Angeletos and La'O (2010) works beyond their specific choice of the information process.

Another line of research our work connects with is to use survey data on forecasts to discipline the expectation formation process. Coibion and Gorodnichenko (2015) and Kohlhas and Walther (2021) test the predictability of aggregate forecasts error, and provide evidence that supports different models with rational expectations and dispersed information. Bordalo, Gennaioli, Ma, and Shleifer (2020), Broer and Kohlhas (2019), and Fuhrer (2018) instead propose various deviations from rationality to account for salient patterns in individual forecasts. The method in our paper can be combined with different types of bounded rationality and helps to explore their general equilibrium implications, as demonstrated in Angeletos, Huo, and Sastry (2021).

Finally, our work is also complementary to the literature that solves models with endogenous information numerically. To maintain the number of state variables to be finite, one common approach is to keep track of a finite number of signals (Townsend, 1983; Hellwig, 2002; Lorenzoni, 2009; Venkateswaran, 2014). Nimark (2017) and Melosi (2014) instead approximate the equilibrium outcome with a finite number of higher-order expectations. Sargent (1991) uses a lower-order ARMA process to approximate the true equilibrium process. Han, Tan, and Wu (2019) solves the model in the frequency domain with the Fourier transformation, and they show that the method in obtaining the fundamental representation proposed in this paper is useful in computing quantitative models. Chiang (2022) extends the analysis beyond linear models to accommodate the effects of dispersed information on higher-order moments. Our numerical strategy differs from the existing methods by solving an exogenous-information equilibrium in each iteration, which helps reduce the number of state variables and increase the convergence speed.

2. An Illustrative Example

In this section, we present a relatively simple beauty-contest model similar to the one considered in Morris and Shin (2002) and Woodford (2003). In this model, higher-order expectations play an important role in shaping aggregate outcomes and the infinite regress problem naturally arises. We use this model to illustrate how our method works.

[6] See Angeletos and Lian (2016) for a more comprehensive review of the literature.

Consider an economy with a continuum of agents. Agent i's best response in period t, a_{it}, is a weighted average of her forecast of an exogenous fundamental, ξ_t, and the aggregate outcome a_t

$$a_{it} = (1-\alpha)\mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[a_t], \qquad \text{where} \qquad a_t = \int a_{it}. \tag{2.1}$$

The parameter $\alpha \in (-1, 1)$ determines the degree of strategic complementarity ($\alpha > 0$) or substitutability ($\alpha < 0$) between agents' actions.[7] The operator $\mathbb{E}_{it}[\cdot]$ denotes the expectation conditional on agent i's information set which will be specified shortly.

We assume that the fundamental ξ_t follows AR(1) process

$$\xi_t = \rho\xi_{t-1} + \eta_t, \quad \eta_t \sim \mathcal{N}(0, 1),$$

and in each period, agents receive a private signal about the fundamental

$$x_{it} = \xi_t + u_{it}, \quad u_{it} \sim \mathcal{N}(0, \sigma^2). \tag{2.2}$$

The information set of agent i contains all the signals up to time t, $\mathcal{I}_{it} = \{x_{it}, x_{it-1}, \ldots\}$. We purposely choose a relatively simple fundamental and signal process which is sufficient to illustrate the key idea of our method, and the analysis will be extended to allow for a much more general information structures in Section 3. Also note that the information here is exogenous, in the sense that its informativeness does not depend on endogenous objects determined in equilibrium. We will discuss the case with endogenous information later in Section 5.

Complete Information Benchmark. In this economy, the variance of the idiosyncratic noise, σ^2, determines the degree of information frictions. Suppose, momentarily, agents observe the fundamental perfectly ($\sigma = 0$), and this fact is common knowledge. It is then immediate that we return to the representative-agent case and the outcome is pinned down solely by the fundamental

$$a_t^* = \xi_t.$$

Furthermore, the strategic complementarity α is irrelevant in determining the equilibrium outcome.

Incomplete Information and Higher-Order Expectations. When $\sigma > 0$, the fundamental can no longer be observed perfectly and agents need to solve a signal extraction problem to infer the fundamental, which represents first-order uncertainty. More importantly, information is dispersed and there is a lack of common knowledge. To infer others' actions, an individual agent also needs to infer other agents' beliefs, and other agents' beliefs about other agents' beliefs, and so on, which represents higher-order uncertainty.

[7] With $\alpha > 1$, there could be multiple equilibria if the action is bounded. By assuming $\alpha \in (-1, 1)$, we can guarantee the existence of a unique equilibrium that can be represented by the sum of infinite higher-order expectations, which satisfies the 'global stability under uncertainty' condition provided by Weinstein and Yildiz (2007).

In fact, the aggregate outcome can be expressed as a function of higher-order expectations. By aggregation, condition (2.1) becomes

$$a_t = (1-\alpha)\overline{\mathbb{E}}_t[\xi_t] + \alpha\overline{\mathbb{E}}_t[a_t], \tag{2.3}$$

where $\overline{\mathbb{E}}_t[\cdot]$ stands for the average expectation in the cross-section of the population. Iterating the above condition once, we have

$$a_t = (1-\alpha)\overline{\mathbb{E}}_t[\xi_t] + (1-\alpha)\alpha\overline{\mathbb{E}}_t\left[\overline{\mathbb{E}}_t[\xi_t]\right] + \alpha^2\overline{\mathbb{E}}_t\left[\overline{\mathbb{E}}_t[a_t]\right],$$

in which the dependence of the aggregate outcome on the second-order expectation appears. By repeatedly iterating condition (2.3), the aggregate outcome can be expressed as a function of the infinite hierarchy of expectations about the fundamental

$$a_t = (1-\alpha)\sum_{k=0}^{\infty}\alpha^k\overline{\mathbb{E}}_t^{k+1}[\xi_t], \tag{2.4}$$

where the higher-order expectation is defined recursively as $\overline{\mathbb{E}}_t^{k+1}[X] \equiv \overline{\mathbb{E}}_t\left[\overline{\mathbb{E}}_t^{k}[X]\right]$.

This higher-order expectation representation remains to be true regardless of the information structure. With complete information, the law of iterated expectations applies and all higher-order expectations are identical to the first-order expectation, in which case $a_t = \overline{\mathbb{E}}_t[\xi_t]$. In contrast, when information is incomplete, higher-order expectations differ from first-order expectations and the equilibrium outcome inherits the properties of all these different expectations. Meanwhile, with dynamic information, the laws of motion of higher-order expectations become increasingly complex as the order increases, which amounts to a computational challenge.

Though expressing the aggregate outcome in terms of higher-order expectations can be helpful for economists to understand the effects of incomplete information,[8] it is not necessary for agents in the economy to compute them when choosing the best action. Similar to the case with perfect information, it is sufficient for agents to obtain the law of motion of the aggregate outcome at the fixed point, and they may bypass the computation of higher-order expectations. In fact, this is the approach that will be taken in this paper.

To solve the equilibrium with incomplete information, the difficulty lies in identifying the right state variables that summarize the past information. In standard complete-information models, it is typically straightforward to find the state variables, such as capital and TFP in real-business-cycle models. In contrast, with dispersed information, the entire history of signals is potentially relevant, and it is not even clear whether there exists such finite-dimensional state variables or not.[9] In what

[8]See Morris and Shin (2002), Woodford (2003), Angeletos and Lian (2018) for example.

[9]In Woodford (2003) and Angeletos and La'O (2010), a guess-and-verify approach is used to solve this type of problem by conjecturing a particular law of motion for the aggregate outcome. However, it remains unclear whether a finite-state law of motion exists or not in general, and if so, what form it takes.

follows, we will explain in details how to overcome this difficulty.

Inference. To provide a road map, we start with the well-known Kalman filter to obtain the forecast rule for the fundamental. We then derive the fundamental representation of the signal process based on the previous forecast rule, which in turn facilitates the use of Wiener filter to obtain the forecast rule for the aggregate outcome. The equilibrium is obtained by solving a fixed point problem in the end.

By the Kalman filter, the first-order expectation about the fundamental ξ_t is given by

$$\mathbb{E}_{it}[\xi_t] = \mathbb{E}_{i,t-1}[\xi_{t-1}] + \left(1 - \frac{\lambda}{\rho}\right)(x_{it} - \mathbb{E}_{i,t-1}[\xi_t]) = \left(1 - \frac{\lambda}{\rho}\right)\frac{1}{1-\lambda L}x_{it}. \tag{2.5}$$

Condition (2.5) is the simply the optimal Bayesian updating: the forecast is a weighted average between the prior mean and the new signal. The weight on the new signal, $1 - \frac{\lambda}{\rho}$, is the familiar Kalman gain where

$$\lambda = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{1}{\rho\sigma^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{1}{\rho\sigma^2}\right)^2 - 4}\right]. \tag{2.6}$$

However, the Kalman filter cannot be directly used to forecast the aggregate outcome a_t. Due to the linear-Gaussian framework, the optimal action a_{it} is a linear function of current and past signals. Denote the policy function as $a_{it} = h(L)x_{it}$, where $h(L) = \sum_{k=0}^{\infty} h_k L^k$. As idiosyncratic shocks wash out in aggregate, the law of motion of the aggregate outcome is then $a_t = h(L)\xi_t = \frac{h(L)}{1-\rho L}\eta_t$. A prerequisite for applying the Kalman filter is that the law of motion of a_t is known ex ante, but $h(L)$ is the equilibrium object to be solved for. This constraint makes us turn to the Wiener filter, with which the forecast of a variable does not require the exact form of its law of motion. This property is particularly useful for the problem at hand.

A key step when applying the Wiener filter is to obtain a fundamental representation of the signal process, $x_{it} = B(L)w_{it}$, where $B(L)$ is invertible and w_{it} is some serially uncorrelated innovation. This is an alternative representation of the original signal process with the property that the history of signals x_i^t and the history of shocks w_i^t contain the same amount of information. Recall that with the original signal representation (2.2), there are two shocks (η_t and u_{it}), but only one signal. Therefore, the underlying shocks contain strictly more information than the signals. This new representation is necessary because the linear projection is ultimately on the space spanned by shocks, which requires that the signals span exactly the same space.

To construct such a fundamental representation, we need to revisit the Kalman filter. From the formula (2.5), the forecast error of the future signal is

$$w_{it+1} \equiv x_{it+1} - \mathbb{E}_{it}[x_{it+1}] = x_{it+1} - \rho\mathbb{E}_{it}[\xi_t] = \frac{1-\rho L}{1-\lambda L}x_{it+1}. \tag{2.7}$$

It follows that the signal can be expressed as a combination of the forecast errors

$$x_{it} = B(L)w_{it} \equiv \frac{1-\lambda L}{1-\rho L} w_{it}, \quad w_{it} \sim \mathcal{N}\left(0, \frac{\rho\sigma^2}{\lambda}\right). \tag{2.8}$$

Equation (2.8) is a fundamental representation and w_{it} is the corresponding fundamental innovation. To see that w_{it} is serially uncorrelated, note that the forecast error w_{it} is orthogonal to the past signals $\{x_{it-1}, x_{it-2}, \ldots\}$ and therefore is uncorrelated with its own past values $\{w_{it-1}, w_{it-2}, \ldots\}$. Meanwhile, $B(L)$ is invertible and the fundamental innovations w_{it} can also be expressed as a function of current and past signals, $w_{it} = B(L)^{-1}x_{it}$. Hence, the signals and the fundamental innovations contain the same amount of information.

Now we are ready to spell out the forecast about a_t. By the Wiener-Hopf prediction formula,[10]

$$\mathbb{E}_{it}[a_t] = \underbrace{\left[\begin{bmatrix} \frac{h(L)}{1-\rho L} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{1-\rho L^{-1}} & \sigma \end{bmatrix}'\right.}_{\text{Cov}(a_t, x_{it})} \underbrace{\left. B\left(L^{-1}\right)^{-1} \frac{\lambda}{\rho\sigma^2}\right]_+ B(L)^{-1}}_{\text{V}(x_{it})^{-1}} x_{it}, \tag{2.9}$$

$$= \frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L)L - h(\lambda)\lambda\frac{1-\rho L}{1-\rho\lambda}\right)x_{it} \tag{2.10}$$

The forecasting formula (2.9) is reminiscent of an OLS estimator. The first term corresponds to the covariance between a_t and the signal, and the second term corresponds to the inverse of the variance of the signal.[11] Different from the standard OLS estimator, the forecast is conditional on information up to time t, excluding the use of the signals realized in the future. The truncation of the sample is achieved by the annihilation operator, $+$, which eliminates L with negative powers. Note that this step is valid only if the last term, $B(L)^{-1}x_{it}$, is uncorrelated over time so that the best forecast of its future values is zero. This is indeed the case as by construction, $B(L)^{-1}x_{it} = w_{it}$, which is serially uncorrelated forecast errors, and it explains why the fundamental representation is needed.

To obtain the forecast formula (2.10), it does not require the particular law of motion of $h(L)$. This allows us to proceed without specifying the state variables, and solve for $h(L)$ directly instead of a guess-and-verify approach. Notice that a new constant $h(\lambda)$ appears, the value of which remains unknown. It turns out that this constant plays an important role in determining the existence and uniqueness of the solution.

[10]Note that in terms of the underlying shocks,

$$a_t = \begin{bmatrix} \frac{h(L)}{1-\rho L} & 0 \end{bmatrix} \begin{bmatrix} \eta_t & u_{it} \end{bmatrix}', \qquad \text{and} \qquad x_{it} = \begin{bmatrix} \frac{1}{1-\rho L} & 1 \end{bmatrix} \begin{bmatrix} \eta_t & u_{it} \end{bmatrix}'.$$

[11]To be precise, $\begin{bmatrix} \frac{h(L)}{1-\rho L} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{1-\rho L^{-1}} & \sigma \end{bmatrix}'$ is a cross-covariance generating function between a_t and x_{it} and $\frac{\lambda}{\rho\sigma^2}B(L)B\left(L^{-1}\right)$ is an auto-covariance generating function of x_{it}.

Fixed Point Problem. Now we solve for the policy function $h(L)$. Using the best response function (2.1) and the forecast formulas (2.5) and (2.10), it follows that

$$h(L)x_{it} = (1-\alpha)\left[\left(1-\frac{\lambda}{\rho}\right)\frac{1}{1-\lambda L}x_{it}\right] + \alpha\left[\frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L)L - h(\lambda)\lambda\frac{1-\rho L}{1-\rho\lambda}\right)x_{it}\right].$$

This condition needs to hold for all possible realizations of the signal x_{it}. After combining like terms, we have

$$\left((L-\lambda)(1-\lambda L) - \frac{\alpha\lambda}{\rho\sigma^2}L\right)h(L) = (1-\alpha)\left(1-\frac{\lambda}{\rho}\right)(L-\lambda) - \alpha\frac{\lambda^2}{\rho\sigma^2(1-\rho\lambda)}h(\lambda)(1-\rho L). \tag{2.11}$$

In condition (2.11), the constant $h(\lambda)$ remains to be determined. There is a continuum of potential solutions to $h(L)$ indexed by the choice of $h(\lambda)$. Meanwhile, note that the term $\left((L-\lambda)(1-\lambda L) - \frac{\alpha\lambda}{\rho\sigma^2}L\right)$ on the left-hand side is a second-order polynomial in L with two roots $\vartheta \in (0,1)$ and ϑ^{-1}

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

To make sure $h(L)$ is an analytic function without any pole inside the unit circle,[12] the constant $h(\lambda)$ has to be set such that ϑ is a root of the right-hand side of equation (2.11) as well, that is

$$(1-\alpha)\left(1-\frac{\lambda}{\rho}\right)(\vartheta-\lambda) - \alpha\frac{\lambda^2}{\rho\sigma^2(1-\rho\lambda)}h(\lambda)(1-\rho\vartheta) = 0.$$

There exists a unique $h(\lambda)$ satisfying this condition, which can then be substituted into condition (2.11) to yield the policy function and the the law of motion of a_t[13]

$$h(L) = \left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}, \qquad \text{and} \qquad a_t = \vartheta a_{t-1} + \left(1-\frac{\vartheta}{\rho}\right)\xi_t. \tag{2.12}$$

Now it is self-evident that the right state variables for economists to keep track of the evolution of the aggregate outcome a_t are simply, (a_{t-1}, ξ_t), the current fundamental and the outcome in the last period. The effects of all the higher-order expectations on the equilibrium outcome are therefore incorporated in the variable ϑ, which we explore further in Section 4.1.

This example provides an elementary illustration of the method in solving models with dispersed information. The key step is to connect the Kalman filter with the Wiener filter in forecasting the endogenous aggregate outcomes, which yields a finite-state representation of the equilibrium and overcomes the difficulty in keeping track of all the higher-order expectations.

[12]With any pole inside the unit circle, the policy rule requires to use of future signals, which is inconsistent with agents' information constraint. For example, $\frac{1}{1-\vartheta^{-1}L}x_{it} = x_{it} + \vartheta x_{i,t+1} + \vartheta^2 x_{i,t+2} + \dots$.

[13]Rondina (2008) obtains a similar analytical solution for models with independent value best response.

3. Rational Expectations Models with Dispersed Information

In this section, we extend the basic idea to models with much more general information and payoff structures. We provide the formula for equilibrium policy rule and characterize the equilibrium properties when informational friction and strategic interaction are jointly present.

3.1 Setup

We restrict our attention to models in which all the variables depend on the underlying Gaussian shocks in a linear way. The input of the model includes two parts: the signal process and a system of equations describing the conditions which the variables need to satisfy. There are three types of variables involved here: an individual agent's own actions, the actions chosen by other agents, and some exogenous fundamentals.

Best Response. In each period t, individual agent i chooses r different actions,

$$\boldsymbol{a}_{it} \equiv \begin{bmatrix} a_{it}^1 & \dots & a_{it}^r \end{bmatrix}'.$$

The best response is

$$\boldsymbol{a}_{it} = \mathbb{E}_{it}[\boldsymbol{\xi}_{it}] + \mathbb{E}_{it}[\boldsymbol{\beta}(L)\boldsymbol{a}_{it}] + \mathbb{E}_{it}[\boldsymbol{\gamma}(L)\boldsymbol{a}_t]. \tag{3.1}$$

The vector $\boldsymbol{\xi}_{it}$ is the vector of exogenous fundamental that may depend on agent i's individual states, and the vector $\boldsymbol{a}_t \equiv \int \boldsymbol{a}_{it}$ is the vector of aggregate outcomes in the economy.

In condition (3.1), we allow $\boldsymbol{\beta}(L)$ and $\boldsymbol{\gamma}(L)$ to be two-sided polynomials in L. For example, if $\gamma(L)$ contains L with negative (positive) power, it implies agents' action depends on future (past) actions of others. In this specification, $\boldsymbol{\beta}(L)$ determines how an agent's action depends on her own future or past actions, which captures the partial equilibrium (PE) considerations. On the other hand, $\boldsymbol{\gamma}(L)$ determines how an agent's action depends on the current, the past, or the future aggregate outcomes in the economy, which captures the general equilibrium (GE) considerations. It nests the commonly used best responses in the literature as special cases, and we further extend it to network games with in Appendix A.7.

Complex Types of Higher-Order Expectations. Similar to the static beauty-contest game in Section 2, higher-order expectations naturally arise with incomplete information. However, the types of higher-order expectations are much richer due to the dynamic nature of strategic complementarities and the across-action dependence in multivariate systems. To further appreciate the richness of various types of higher-order expectations underneath condition (3.1), we look into the details of it in two special cases.

First, consider a univariate best response ($r = 1$) with static and forward-looking complementarity,

$$a_{it} = \mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[a_t] + \beta\mathbb{E}_{it}[a_{it+1}] + \gamma\mathbb{E}_{it}[a_{t+1}], \tag{3.2}$$

where $\beta(L) = \beta L^{-1}, \gamma(L) = \alpha + \gamma L^{-1}$, and $\xi_{it} = \xi_t$. This type of best response nests the incomplete-information version of the dynamic IS curve, the New Keyenesian Philips curve, or the asset pricing equation, as in Allen, Morris, and Shin (2006), Nimark (2008), and Angeletos and Huo (2021), and the commonly used static beauty contests. Let $\zeta_t \equiv \frac{1}{1-\beta L}\xi_t$, the set of relevant higher-order expectations include

$$\overline{\mathbb{E}}_{t_1}[\overline{\mathbb{E}}_{t_2}[\cdots[\overline{\mathbb{E}}_{t_h}[\zeta_{t+k}]\cdots]]],$$

for any $k \geq 0$, $h \in \{2, ..., k\}$, and $\{t_1, t_2, ..., t_h\}$ such that $t = t_1 < t_2 < ... < t_h = t+k$. Considering the higher-order expectations on the future fundamental up to k periods ahead, there are k types of second-order expectations, plus $\frac{k\times(k+3)}{2}$ types of third-order expectations, plus $\frac{(h-2)k(k+1)}{2} + k$ types of h-th order expectations for all $h \leq k$.

Second, consider a multivariate best response ($r > 1$) where agent i's multiple actions are given by

$$\begin{bmatrix} a_{it}^1 \\ \vdots \\ a_{it}^r \end{bmatrix} = \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_r \end{bmatrix} \mathbb{E}_{it}[\xi_t] + \begin{bmatrix} \gamma_{11} & \dots & \gamma_{1r} \\ \vdots & \ddots & \vdots \\ \gamma_{1r} & \dots & \gamma_{rr} \end{bmatrix} \begin{bmatrix} a_t^1 \\ \vdots \\ a_t^r \end{bmatrix}.$$

This type of best response can represent the the incomplete-information NK model where the aggregate demand and aggregate supply blocks interact with each other (Angeletos and Lian, 2018), or an incomplete-information multi-sector production network model (La'O and Tahbaz-Salehi, 2020). In matrix form, the aggregate outcomes can be expressed as

$$\boldsymbol{a}_t = \sum_{k=0}^{\infty} \boldsymbol{\gamma}^k \boldsymbol{\omega} \overline{\mathbb{E}}^{k+1}[\xi_t].$$

As an example, when $\boldsymbol{\omega} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $\boldsymbol{\gamma} = \begin{bmatrix} 0 & \alpha \\ \alpha & 0 \end{bmatrix}$, the relevant higher-order expectations are

$$\left\{\overline{\mathbb{E}}^1[\xi_t], \overline{\mathbb{E}}^3[\xi_t], \overline{\mathbb{E}}^5[\xi_t], \dots\right\}, \quad \text{and} \quad \left\{\overline{\mathbb{E}}^2[\xi_t], \overline{\mathbb{E}}^4[\xi_t], \overline{\mathbb{E}}^6[\xi_t], \dots\right\},$$

for a_t^1 and a_t^2, respectively. As $\boldsymbol{\gamma}^k\boldsymbol{\omega}$ can capture various types of weighted averages of higher-order expectations, the outcomes therefore display much richer dynamics than the single-action case in Section 2.

Information Structure. The process of the fundamental $\boldsymbol{\xi}_{it}$ is specified as

$$\boldsymbol{\xi}_{it} = \boldsymbol{\theta}(L)\boldsymbol{s}_{it} = \frac{\widetilde{\boldsymbol{\theta}}(L)}{\prod_{k=1}^{N_\rho}(1-\rho_k L)} \boldsymbol{s}_{it}, \tag{3.3}$$

The auto-regressive parameters $\{\rho_k\}$ with $|\rho_k| < 1$ determine the persistence of the fundamental and the external propagation of the equilibrium outcomes. The underlying vector of Gaussian shocks $\boldsymbol{s}_{it} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is serially uncorrelated with length m, which contains both common and idiosyncratic components. In our setup, the existence of the idiosyncratic shocks is the root of information incompleteness. The aggregate outcome, on the other hand, only depends on the common shocks as idiosyncratic shocks simply wash out in aggregate.

Each period, instead of observing the fundamental directly, the individual agent i receives a vector of signals about the underlying state of the economy. We denote the stochastic process of the signals as follows

$$\boldsymbol{x}_{it} = \mathbf{M}(L)\boldsymbol{s}_{it}, \tag{3.4}$$

where $\boldsymbol{x}_{it}$ is the vector of signals with length n. In this section, we focus on the exogenous-information economy, in the sense that $\mathbf{M}(L)$ is exogenously specified. In contrast, when signals contain variables that are determined in equilibrium, its information content is endogenous and the structure of $\mathbf{M}(L)$ is part of the equilibrium. However, it is important to note that for an *individual* agent, she always takes the process $\mathbf{M}(L)$ as exogenously given, regardless whether it is determined in equilibrium or not, a point we revisit in Section 5.

With all the elements in the environment specified, it is straightforward to define the equilibrium of this economy.

Definition 1. *Given the exogenous signal process (3.4), a Bayesian-Nash equilibrium is a policy rule $\boldsymbol{a}_{it} = \boldsymbol{h}(L)\boldsymbol{x}_{it}$ that satisfies the best response condition (3.1), and where the aggregate outcome is consistent with individual agents' choice: $\boldsymbol{a}_t = \int \boldsymbol{a}_{it}$.*

As aforementioned, the entire history of signals could be relevant in forecasting the fundamental and the aggregate outcome, and it is not clear ex ante whether there exists a set of sufficient statistics to summarize the history. In this section, we impose the assumption that the signals follow an ARMA process and the primitives in the best response are rational functions of L. As we shall show momentarily, the policy rule in equilibrium inherits this property and only a finite number of state variables are required. As a by-product, the seemingly complex sum of the infinite dynamic higher-order expectations follows a relatively simple process as well.

Assumption 1. *The signal $\boldsymbol{x}_{it}$ follows a finite ARMA process, and all the elements in matrices $\boldsymbol{\beta}(L)$, $\boldsymbol{\gamma}(L)$, and $\boldsymbol{\theta}(L)$ are rational functions of L.*

3.2 Fundamental Representation and Wiener-Hopf prediction formula

Parallel to the analysis in Section 2, we start with inference problems using the Kalman filter. This step helps construct the fundamental representation of the signal process, which builds a bridge to the Wiener filter. Different from Section 2, we now allow a multivariate system, resulting in a more involved procedure.

Given a signal process (3.4), there always exists an alternative representation of the same signal process

$$\boldsymbol{x}_{it} = \mathbf{B}(L)\boldsymbol{w}_{it}, \tag{3.5}$$

such that $\mathbf{B}(L)$ is an invertible,[14] and $\boldsymbol{w}_{it}$ is a vector of serially uncorrelated Gaussian shocks with covariance matrix $\mathbf{V}$ which can be constructed by the history of signals $\boldsymbol{w}_{it} = \mathbf{B}(L)^{-1}\boldsymbol{x}_{it}$. This is the fundamental representation, which shares the same auto-covariance generating function as the original representation

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

The important property of the fundamental representation is that the sequence of the signal $\boldsymbol{x}_i^t$ contains the same amount of information as the sequence of the fundamental innovations $\boldsymbol{w}_i^t$. With this representation, one can apply the following Wiener-Hopf prediction formula.

Wiener-Hopf Prediction Formula. *Let f_t be a univariate co-variance stationary process $f_t = \phi(L)\boldsymbol{s}_{it}$, where $\phi(L) = \sum_{k=-\infty}^{\infty}\phi_k L^k$. The optimal prediction of f_t is given by*

$$\mathbb{E}_{it}[f_t] = \left[\phi(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_+ \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it}. \tag{3.6}$$

Proof. See Appendix A.2 for the proof. □

Notice that the exact law of motion of f_t is not required for the forecasting problem. This is necessary for solving the equilibrium policy rule as the law of motion of the aggregate outcome is not known ex ante.

One the one hand, the Kalman filter is inadequate for the forecasting problem as it requires the exact law of motion ex ante. On the other hand, the Kalman filter helps to construct the fundamental representation Kalman filter, despite the additional complication of a multivariate system. Towards this goal, it is necessary to set up the state-space representation of the signal process.

Lemma 3.1. *Under Assumption 1, the signal process admits a state-space representation given by*

$$\boldsymbol{z}_{it} = \mathbf{F}\boldsymbol{z}_{it-1} + \boldsymbol{\Phi}\boldsymbol{s}_{it}, \quad \textit{and} \quad \boldsymbol{x}_{it} = \mathbf{H}\boldsymbol{z}_{it} + \boldsymbol{\Psi}\boldsymbol{s}_{it}, \tag{3.7}$$

where the eigenvalues of $\mathbf{F}$ all lie inside the unit circle.

The following theorem then provides the desired fundamental representation based on the steady-state Kalman filter.

[14]This is equivalent to that $\mathbf{B}(L)$ is a square matrix and the determinant of $\mathbf{B}(L)$ does not contain any roots within the unit circle.

Fundamental Representation. *Given the state-space representation 3.7, there exist matrices* $\mathbf{P}$ *and* $\mathbf{K}$ *that satisfy*

$$\mathbf{P} = \mathbf{F}[\mathbf{P} - \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1}\mathbf{H}\mathbf{P}]\mathbf{F}' + \mathbf{\Phi}\mathbf{\Phi}', \quad \textit{and} \quad \mathbf{K} = \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1}. \tag{3.8}$$

The fundamental representation is given by

$$\mathbf{B}(L) = \mathbf{I} + \mathbf{H}[\mathbf{I} - \mathbf{F}L]^{-1}\mathbf{F}\mathbf{K}L, \qquad \mathbf{B}(L)^{-1} = \mathbf{I} - \mathbf{H}[\mathbf{I} - (\mathbf{F} - \mathbf{F}\mathbf{K}\mathbf{H})L]^{-1}\mathbf{F}\mathbf{K}L, \tag{3.9}$$

associated with the co-variance matrix $\mathbf{V}$

$$\mathbf{V} = \mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}'. \tag{3.10}$$

Proof. See Chapter 13.5 in Hamilton (1994). □

In equation (3.8), $\mathbf{K}$ is the celebrated Kalman gain matrix. In equation (3.9), the eigenvalues of $\mathbf{F} - \mathbf{F}\mathbf{K}\mathbf{H}$ determine the persistence of prior about the underlying state, which in turn shape the learning dynamics. To see this in a more explicit manner, we unpack formula (3.6) when Assumption 1 holds.

Proposition 3.1. *Under Assumption 1 and assume that* $f_t = \phi(L)\boldsymbol{s}_{it} = \sum_{k=0}^{\infty} \phi_k \boldsymbol{s}_{i,t-k}$. *The optimal prediction of* f_t *is*[15]

$$\begin{aligned}\mathbb{E}_{it}[f_t] =& \phi(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\mathbf{V}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it} - \\ &\left(\sum_{k=1}^{u} \frac{1}{L - \lambda_k} \frac{\phi(\lambda_k)\mathbf{G}(\lambda_k)}{\lambda_k^{v-u} \prod_{\tau \neq k}(\lambda_k - \lambda_\tau)} + \sum_{k=0}^{v-u-1} \frac{1}{k! L^{v-u-k}} \left[\frac{\phi(L)\mathbf{G}(L)}{\prod_{\tau=1}^{u}(L - \lambda_\tau)}\right]_{L=0}^{(k)}\right) \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it}.\end{aligned} \tag{3.11}$$

where $\{\lambda_k\}_{k=1}^{u}$ *are non-zero eigenvalues of* $\mathbf{F} - \mathbf{F}\mathbf{K}\mathbf{H}$ *which lie inside the unit circle,* v *is the dimension of* $\mathbf{F}$*, and* $\mathbf{G}(L)$ *is a polynomial matrix in* L *with finite degree derived in Appendix A.4.*

The component in the first line of (3.11) corresponds to the optimal forecast when both past and future signals are available. The component in the second line of (3.11) is the necessary adjustment due to the annihilation operator, +, when future signals are prohibited. We provide a formula (A.1) implementing the annihilation operator in Appendix A.3. Note that $\mathbf{B}(L)^{-1}$ contains the component $\frac{1}{\prod_{k=1}^{u}(1-\lambda_k L)}$, which implies that the eigenvalues of $\mathbf{F} - \mathbf{F}\mathbf{K}\mathbf{H}$ belong to the AR parameters of $\mathbb{E}_{it}[f_t]$, adding additional persistence due to learning.

An alternative way to construct the fundamental representation is to conduct spectral factorization on the auto-covariance generating function proposed by Rozanov (1967), which is used in Taub (1989), Rondina (2008), Miao, Wu, and Young (2021), and so on. This method requires to remove inside poles of $\mathbf{B}(L)$ by Gaussian elimination and polynomial spectral factorization to make sure $\mathbf{B}(L)$ does not

[15] In equation (3.11), we use $[g(L)]_{L=\delta}^{(k)}$ to denote the k-th derivative of $g(L)$ evaluated at $L = \delta$.

contain negative L in expansion, and to sequentially remove inside roots of $\det[\mathbf{B}(L)]$ by Blaschke factor matrices to make sure that the spectral factorization is canonical (or $\mathbf{B}(L)$ is invertible).[16] Our method has two main advantages: first, the Kalman filter is easy and robust to implement numerically, and therefore is better suited for quantitative analysis; second, the explicit representation from the Kalman filter allows a general closed-form formula (3.11) for the forecasts. This generality facilitates the proof the Theorem 1 and Corollary 3.1 for any stationary ARMA signal process.

3.3 Equilibrium Policy Rule

In this subsection, we builds on the tools developed earlier to solve for the equilibrium policy rule. Supposing individual agents' action is $\boldsymbol{a}_{it} = \boldsymbol{h}(L)x_{it}$, the aggregate outcome can be expressed as

$$\boldsymbol{a}_t = \int \boldsymbol{a}_{it} = \boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\boldsymbol{s}_{it}, \tag{3.12}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix that selects the common shocks, i.e., $\int \boldsymbol{s}_{it} = \boldsymbol{\Lambda}\boldsymbol{s}_{it}$. When $\boldsymbol{\Lambda} \neq \mathbf{I}$, the information is dispersed in the economy.

Parallel to the analysis in Section 2, given a perceived law of motion (3.12), agents can form expectations about the fundamentals, the individual actions, and the aggregate outcome via the forecast rule (3.11). The best response then leads to a functional equation for the equilibrium policy rule. Instead of looking for the sequences of infinite coefficients on how to use the history of signals, we can look for a finite number of analytic functions, as shown in the following proposition.

Lemma 3.2. *If $\boldsymbol{h}(L)$ is an equilibrium policy rule, then it satisfies the following condition*

$$\mathbf{T}(L)\text{vec}[\boldsymbol{h}'(L)] = \mathbf{D}_1(L)\psi + \mathbf{D}_2(L) \tag{3.13}$$

where $\mathbf{T}(L)$ is an $rn \times rn$ matrix given by

$$\mathbf{T}(L) \equiv (\boldsymbol{\beta}(L) - \mathbf{I}) \otimes (\mathbf{M}(L^{-1})\mathbf{M}'(L)) + \boldsymbol{\gamma}(L) \otimes (\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L)), \tag{3.14}$$

$\mathbf{D}_1(L)$ and $\mathbf{D}_2(L)$ are exogenous matrices constructed in Appendix A.5, and $\boldsymbol{\psi}$ is a vector of undetermined constants. Particularly, $\mathbf{D}_1(L)$ is with full column rank N_ψ.

The structure of the system (3.13) resembles that of the pure forecasting problem in (3.11). On the left-hand side, $\mathbf{T}(L)$ captures both the intertemporal dependence on an individual agent's own action and the dynamic coordination with other agents' actions. The right-hand side collects the forecasts of the fundamental and the necessary adjustments due to the annihilation operator. Similar to the second

[16] When $\mathbf{M}(L)$ is a square matrix, one only needs to use the Blaschke factor matrix to flip out the inside roots of $\det[\boldsymbol{\rho}_{xx}(L)]$, see Kasa (2000), Kasa, Walker, and Whiteman (2014), Rondina and Walker (2021), Acharya, Benhabib, and Huo (2021) for example.

line in (3.11), the inference of the endogenous variables yields constants which are linear combinations of $\boldsymbol{h}(\lambda_k)$, $\boldsymbol{h}^{(\tau)}(0)$, and so on. There are N_ψ such endogenous constants that remain to be determined.

Condition (3.13) also helps better understand the role of Assumption 1. Directly, all the elements in $\mathbf{T}(L)$ are rational functions in L by construction. Indirectly, the forecast rule (3.11) reveals that all the elements in $\mathbf{D}_1(L)$ and $\mathbf{D}_2(L)$ are rational functions in L as well. Even though ex-ante the form of $\boldsymbol{h}(L)$ is unknown, to satisfy condition (3.13), it is clear that the elements in $\boldsymbol{h}(L)$ must inherit the property of the primitives and be rational functions in L as well.

To obtain the policy rule, one may attempt to simply invert matrix $\mathbf{T}(L)$ in condition (3.13)

$$\text{vec}[\boldsymbol{h}'(L)] = \frac{\text{adj}[\mathbf{T}(L)]}{\det[\mathbf{T}(L)]}(\mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L)).$$

This step is valid only if all elements of the policy rule $\boldsymbol{h}(L)$ is an analytic function in L which does not contain any pole inside the unit circle, that is, agents only use current or past signals. The set of constants $\boldsymbol{\psi}$ can be used to remove the inside roots of $\det[\mathbf{T}(L)]$.[17] As a result, the existence and the uniqueness of the equilibrium hinges on the number of free constants versus the number of the inside roots of $\det[\mathbf{T}(L)]$.

Among all the roots of $\det[\mathbf{T}(L)]$, we denote

- $\{\zeta_1, \dots, \zeta_{N_\zeta}\}$ as the N_ζ roots that lie inside the unit circle, and
- $\{\vartheta_1^{-1}, \dots, \vartheta_{N_\vartheta}^{-1}\}$ as the N_ϑ roots that lie outside the unit circle.

The equilibrium policy rule is then given below.

Theorem 1 (Solution)**.** *Generically, there exists a unique equilibrium iff $N_\psi = N_\zeta$, which is given by*

$$\text{vec}[\boldsymbol{h}(L)] = \mathbf{T}(L)^{-1}(\mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L)),$$

where $\boldsymbol{\psi}$ satisfies the condition that for $i \in \{1, \dots, rn\}$ and $j \in \{1, 2, \dots, N_\psi\}$,

$$\det\begin{bmatrix}\mathbf{T}_1(\zeta_j) & \dots & \mathbf{T}_{i-1}(\zeta_j) & \mathbf{D}_1(\zeta_j)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_j) & \mathbf{T}_{i+1}(\zeta_j) & \dots & \mathbf{T}_{rn}(\zeta_j)\end{bmatrix} = 0.$$

Proof. See Appendix A.6 for the construction of $\boldsymbol{\psi}$ and a detailed description of the condition for equilibrium existence and uniqueness. □

The condition for a uniqueness equilibrium is reminiscent of the one in Whiteman (1983), and we generalize the model environment to incorporate dispersed information and coordination. The main difficulty in solving the problem in the time domain is to identify the right state variables that summarize the relevant history. By applying the Wiener filter in the frequency domain, one does not need to identify the state variables ex-ante and the task is transformed into solving for a particular

[17] As explained in Appendix A.5, the poles of $\boldsymbol{h}(L)$ cannot come from $\mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L)$.

analytic function. It turns out that this problem remains tractable, and the dependence on the history is encoded in $\{\vartheta_1, \ldots, \vartheta_{N_\vartheta}\}$.

Extension to Network Games. We conclude this subsection by pointing out an important extension. So far, we have focused on environments in which agents' actions differ from each other only due to the realizations of idiosyncratic shocks. However, the method we have developed and the main theoretical results easily extend to more complicated models where agents also differ in payoff structures and the information structures. For example, firms in different industries are interconnected through supply chains, but they may not share the same information about the TFP growth in a particular sector; savers' and borrowers' expenditures both depend on aggregate demand and real interest rates, but they may have different expectations about the inflation rates. Effectively, our method can be applied to these types of network games with dispersed information, given that Assumption 1 holds for all agents. In Appendix A.7, we show how to construct the counterpart of condition (3.13) in such network games with a richer types of heterogeneities.[18]

3.4 Implications

In this subsection, we discuss further the implications of our results on the information incompleteness, the GE feedback effects, and the dynamic properties of the equilibrium outcomes.

Finite-State Representation. A direct implication of Theorem 1 is that the equilibrium outcomes admit a finite-state recursive representation.

Corollary 3.1 (Finite-state representation)**.** *The equilibrium outcome permits a finite ARMA representation*

$$\boldsymbol{a}_{it} = \frac{\boldsymbol{a}(L)}{\prod_{k=1}^{N_\vartheta}(1-\vartheta_k L)\cdot\prod_{k=1}^{N_\rho}(1-\rho_k L)}\boldsymbol{s}_{it} \tag{3.15}$$

where $\boldsymbol{a}(L)$ is a lag polynomial matrix with a finite degree, $\{\vartheta_k\}$ is the vector of endogenous coefficients, and $\{\rho_k\}$ is the vector of exogenous coefficients from (3.3).

This result highlights that when signals follow finite ARMA processes, the equilibrium outcome inherits this property. In contrast to the conventional wisdom that it is necessary to keep track of the entire history of signals when information is dispersed and persistent (Townsend, 1983), Corollary 3.1 instead shows that a finite number of statistics are sufficient to summarize the history in equilibrium.

The propagation dynamics are determined by two set of parameters: first, $\{\rho_k\}$ are the AR parameters of the exogenous fundamental, which can be viewed as the external propagation mechanism; second, $\{\vartheta_k\}$ are determined in equilibrium, which can be viewed as the endogenous propagation mechanism. It is important to note that the parameters $\{\lambda_k\}$ that determine the persistence of forecasts in (3.11) do *not* enter the equilibrium outcome, though they show up for each of the higher-order

[18] Angeletos and Huo (2021) utilize this result in the context of a HANK model with forward-looking complementarities.

expectations. The seemingly magical cancellation of $\{\lambda_k\}$ makes a finite-state representation possible and reduces the dimension of state variables, a property we shall revisit in Section 5.2.

As a byproduct, Corollary 3.1 shows that the guess-and-verify approach in Woodford (2003) works beyond the particular model environment. Furthermore, it helps researchers to determine the right law of motion to conjecture if they pursue the guess-and-verify approach for more general model economies.

Role of Incomplete Information and GE. Equation (3.14) summarizes how the PE consideration, $\boldsymbol{\beta}(L)$, the GE consideration, $\boldsymbol{\gamma}(L)$, and the informational friction, $\mathbf{M}(L)$, jointly shape the roots $\{\vartheta_1, \ldots, \vartheta_{N_\vartheta}\}$. The following result establishes the necessary condition for their interactive effects.

Corollary 3.2 (Incompleteness and GE)**.** *The persistence of equilibrium outcome depends jointly on the informational friction and the payoff structure only if: (1) information is incomplete, $\boldsymbol{\Lambda} \neq \mathbf{I}$, and (2) the GE consideration is present, $\boldsymbol{\gamma}(L) \neq \mathbf{0}$.*

To better understand this result, consider the following special cases. First, suppose that agents have perfect information, which corresponds to the frictionless case with a representative agent. In this case, agents can observe all the shocks, or $\mathbf{M}(L) = \mathbf{I}$, which implies

$$\det[\mathbf{T}(L)] = \det[\boldsymbol{\beta}(L) + \boldsymbol{\gamma}(L) - \mathbf{I}].$$

Therefore, only the payoff structure matters for the persistence, such as the magnitude of adjustment costs or consumption habit.

Secondly, suppose that agents have common information (not necessarily perfect), i.e., $\boldsymbol{\Lambda} = \mathbf{I}$, an assumption imposed by most DSGE literature. In this case, all agents are identical to each other. Similar to the first case, the distinction between PE and GE becomes irrelevant, and only the composite effects $\boldsymbol{\beta}(L) + \boldsymbol{\gamma}(L)$ enters the determinant of the matrix $\mathbf{T}(L)$

$$\det[\mathbf{T}(L)] = \det[\boldsymbol{\beta}(L) + \boldsymbol{\gamma}(L) - \mathbf{I}] \cdot \det\left[\mathbf{M}(L^{-1})\mathbf{M}'(L)\right].$$

Clearly, in this case, the roots of $\det[\mathbf{T}(L)]$ are determined *separately* by the payoff structure and the informational friction.

Thirdly, suppose the coordination motive or the GE feedback effect is muted, $\boldsymbol{\gamma}(L) = \mathbf{0}$, but information may still be dispersed. In this case, agents only care about their own fundamentals, and whether the information is private or not is irrelevant. Particularly, the determinant of the matrix $\mathbf{T}(L)$ becomes

$$\det[\mathbf{T}(L)] = \det[\boldsymbol{\beta}(L) - \mathbf{I}] \cdot \det\left[\mathbf{M}(L^{-1})\mathbf{M}'(L)\right].$$

Corollary 3.2 underscores the importance of higher-order expectations in modifying the equilibrium behavior. In the absence of either information incompleteness or GE considerations, only first-order uncertainty matters for equilibrium outcomes, as in the aforementioned special cases. The interactive effects take place exactly when higher-order uncertainty is present.

Dimension Reduction of Higher-Order Expectations. The last implication can be viewed independent of any equilibrium concept, but only as a property of the linear projection. Essentially, the weighted average of infinite higher-order expectations can be much simpler than it appears to be.

We impose the following condition on the primitives in the best response, which helps guarantee the existence of a uniqueness equilibrium.

Assumption 2. *Agents are forward-looking, that is,* $\boldsymbol{\beta}(L) = \sum_{k=0}^{\infty} \boldsymbol{\beta}_k L^{-k}$ *and* $\boldsymbol{\gamma}(L) = \sum_{k=0}^{\infty} \boldsymbol{\gamma}_k L^{-k}$. *In addition,* $\det[\mathbf{I} - \boldsymbol{\beta}(L)]$ *only has inside roots, and all eigenvalues of* $\mathbf{I} - \boldsymbol{\beta}(1) - \boldsymbol{\gamma}(1)$ *are inside the unit circle.*

Corollary 3.3 (HOE)**.** *Under Assumption 1 and 2, the infinite sum of higher-order expectations follows a finite ARMA process*

$$\boldsymbol{a}_t = \sum_{k=0}^{\infty} \overline{\mathbb{F}}_t^k [(\mathbf{I} - \boldsymbol{\beta}(L))^{-1} \boldsymbol{\xi}_{it}] = \frac{\boldsymbol{a}(L)}{\prod_{k=1}^{N_\vartheta}(1 - \vartheta_k L) \cdot \prod_{k=1}^{N_\rho}(1 - \rho_k L)} \boldsymbol{\Lambda} \boldsymbol{s}_{it},$$

where $\overline{\mathbb{F}}^{k+1}[X] = \overline{\mathbb{E}}[(\mathbf{I} - \boldsymbol{\beta}(L))^{-1} \boldsymbol{\gamma}(L) \overline{\mathbb{F}}^k [X]]$.

As already mentioned, the complexity of higher-order expectations is increasing with its order in the sense that more state variables are required to describe their laws of motion. To compute each of the infinite higher-order expectations independently, an infinite number of state variables are needed. Corollary 3.3 shows that the infinite sum of these higher-order expectations magically reduces to a much simpler object that follows a finite ARMA process. In a special case where the strategic complementarity is static, i.e., $\boldsymbol{\gamma}(L) = \boldsymbol{\gamma}$, this equivalence takes a particularly sharp form that the sum of higher-order expectations is the same as its corresponding first-order expectation with more noisy signals, as shown in Huo and Pedroni (2020).

This equivalence also helps reconcile two different perspectives on the rational expectations equilibrium with dispersed information. On the one hand, an agent in the economy only needs to know the law of motion of the aggregate outcome. On the other hand, economists may find it more informative to think about all the higher-order expectations, which is much more sophisticated than what agents need to make their decisions. Corollary 3.3 connects the two approaches by showing the apparent complexity in the latter approach reduces by a large extent at the fixed point.

4. Applications

In this section, we demonstrate how our method developed in Section 3 can help obtain applied lessons. We start by showing that the finite-state representation makes it possible to derive closed-form solutions and prove comparative statics. We then contrast our solution under rational expectations with an alternative approach in dealing with heterogeneous beliefs both in terms of allocation and the properties on forecasts. We also utilize the generality of the method to incorporate belief distortions into models with incomplete information, which allows one to explore their general equilibrium implications.

4.1 Deriving Comparative Statics

In this subsection, we revisit the beauty-contest model (2.1) introduced in Section 2

$$a_{it} = (1-\alpha)\mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[a_t],$$

and extend it to allow agents to observe a public signal (x_{it}^1) in addition to the private signal (x_{it}^2)

$$x_{it}^1 = \xi_t + \varepsilon_t, \quad \text{and} \qquad x_{it}^2 = \xi_t + u_{it}.$$

where $\varepsilon_t \sim \mathcal{N}(0, \tau_\varepsilon^{-1})$ represents the common noise and $u_{it} \sim \mathcal{N}(0, \tau_u^{-1})$ is the private noise.[19] In a similar framework, Lorenzoni (2009) and Angeletos and La'O (2010) solve the model numerically and highlight the response to the common noise. This generates variations in the aggregate outcome that are independent of the fundamental, which can be interpreted as animal spirits or sentiments. In our setup, applying Theorem 1 leads to the following characterization of the equilibrium in closed form.

Proposition 4.1. *The aggregate outcome is given by*

$$a_t = \left(1 - \frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\xi_t + \frac{\tau_\varepsilon}{\tau_\varepsilon + (1-\alpha)\tau_u}\left(1 - \frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\varepsilon_t \tag{4.1}$$

where the persistence $\vartheta \in [0, \rho]$ is given by

$$\vartheta = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{\tau_\varepsilon + (1-\alpha)\tau_u}{\rho}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\tau_\varepsilon + (1-\alpha)\tau_u}{\rho}\right)^2 - 4}\right]. \tag{4.2}$$

Relative to the perfect information benchmark, $a_t^* = \xi_t$, the incomplete-information version (4.1) modifies it in the following way: (1) the response to the fundamental shock displays a dampened impact effect and a more gradual build up, both of which are governed by ϑ. (2) Besides the fundamental shock, the common noise also contributes to aggregate fluctuations. Proposition 4.1 immediately reveals that the fluctuations due to the common noise share the same persistence ϑ, but the impact response depends on the amount of information in the public domain relative to that in the private domain.

The key variable that determines the dynamics of the aggregate outcome is ϑ, which corresponds to the reciprocal of the outside root of $\det[\mathbf{T}(L)]$. Note that ϑ depends on all the structural parame-

[19]In Appendix A.11, we describe in details how to map the primitives in this model into the general framework outlined in Section 3.

ters, and it summarizes the interaction between incomplete information and GE consideration. For example, the precision of the private signal, τ_u, enters ϑ with an additional discounting according to $1-\alpha$. The coordination motive makes the private signal less useful in inferring the aggregate outcome compared with the public signal.

The closed-form solution facilitates a transparent comparative statics analysis. The following proposition illustrates how the persistence and volatility of the aggregate outcome vary with informational friction and GE consideration.

Proposition 4.2. 1. *The endogenous persistence ϑ is increasing in α.*

2. *The endogenous persistence ϑ is decreasing in τ_u and τ_ε. Furthermore, changes in τ_ε have a larger (smaller) impact on ϑ than τ_u when $\alpha > 0 (< 0)$*

$$\frac{\partial \vartheta}{\partial \tau_u} = (1-\alpha)\frac{\partial \vartheta}{\partial \tau_\varepsilon}.$$

3. *The volatility of aggregate outcome driven by the common noise, $\mathbb{V}[a_t|\xi_t]$, is increasing in α, while that driven by the fundamental, $\mathbb{V}[a_t|\varepsilon_t]$, is decreasing in α.*

Part 1 of Proposition 4.2 shows that fixing the informational friction, the endogenous persistence is increasing in the GE consideration α. Woodford (2003) emphasizes that higher-order expectations respond more sluggishly compared with first-order expectations, and the aggregate outcome may display a hump-shaped response when the reliance on the former is sufficiently strong. This additional inertia is exactly captured by the term $\frac{1}{1-\vartheta L}$ in condition (4.1).

Part 2 of Proposition 4.2 states that the endogenous persistence is amplified with a higher degree of informational friction. However, changing the informational friction in the public domain versus that in the private domain have differential impacts on ϑ, as the relative dependence on the two signals is shaped by the need to be in line with others.

The last part of Proposition 4.2 looks into the conditional volatilities. As captured by $\frac{\tau_\varepsilon}{\tau_\varepsilon+(1-\alpha)\tau_u}$, a larger α leads to more intensive use of public signal, and therefore a larger loading on the common noise. At the same time, as captured by $1-\frac{\vartheta}{\rho}$, a larger α also leads to more weight on higher-order expectations and a more dampened response overall. Despite the presence of competing forces, the analytical result allows us to prove that the volatility conditional on the fundamental (the noise-driven fluctuations) is always increasing in α, and the volatility conditional on the common noise (the fundamental-driven fluctuations) is always decreasing in α.

4.2 HANK Model with Heterogeneous Information Structures

Beyond the univariate static beauty contest model, the analysis can be extended to network games with incomplete information. Particularly, we focus on a HANK type model with incomplete information

(Angeletos and Huo, 2021; Auclert, Rognlie, and Straub, 2020) and explore how heterogeneities in MPC, income exposure, and informational friction interact with each other.

Following Angeletos and Huo (2021), we consider a perpetual-youth, overlapping-generations version of the HANK model. The perceived different mortality risks map to different MPCs.[20] Suppose that there are two groups of consumers indexed by $g \in \{1, 2\}$ with mass π_g, where group 1 stands for the high MPC group and group 2 stands for the low MPC group. We denote the MPC as m_g and the discount factor as $1 - m_g$. In addition, different groups can have different exposures to the business cycle (Patterson, 2019): the (log) income of group g is $y_t^g = \phi_g y_t$, where ϕ_g captures the group specific income exposure to aggregate output.[21].

The dynamics of the average consumption in group g can be expressed as

$$c_{g,t} = -(1 - m_g) \sum_{k=0}^{\infty} (1 - m_g)^k \overline{\mathbb{E}}_{g,t}[r_{t+k}] + m_g \phi_g \sum_{k=0}^{\infty} (1 - m_g)^k \overline{\mathbb{E}}_{g,t}[y_{t+k}], \tag{4.3}$$

where $\overline{\mathbb{E}}_{g,t}$ stands for the average expectation within group g, and the aggregate output follows $y_t = \sum_g \pi_g c_{g,t}$. Condition (4.3) can be viewed as a version of the Permanent Income Hypothesis. The consumption is a function of the present discounted value of income, incorporating variations in the real interest rate and incomplete information. Note that it is the product of the MPC and the income exposure that determines the strength of the general equilibrium consideration. Also note that condition (4.3) together with the aggregate output effectively consists of a forward-looking network game.

Assume that the central bank directly controls the real interest rate r_t which follows an AR(1) process, and an individual consumer i in group g observes a noisy signal $x_{i,g,t}$ about r_t every period

$$r_t = \rho r_{t-1} + \eta_t, \qquad x_{i,g,t} = r_t + u_{i,g,t},$$

where $u_{i,g,t} \sim \mathcal{N}(0, \tau_g^{-1})$. Importantly, we allow different groups to face heterogeneous information structures indexed by the signal precision τ_g. This is motivated by the empirical evidence that the informational frictions faced by consumers depend on their socioeconomic status (Broer, Kohlhas, Mitman, and Schlafmann, 2021; Rozsypal and Schlafmann, 2022). Different levels of τ_g in our environment captures such dependence. Thus, as in section 2, the first-order expectation about r_t of an individual consumer i in group g is

$$\mathbb{E}_{i,g,t}[r_t] = \left(1 - \frac{\lambda_g}{\rho}\right) \frac{1}{1 - \lambda_g L} x_{i,g,t},$$

where λ_g captures the persistence of the first-order belief and could depend on groups.

[20]Following Piergallini (2007), Del Negro, Giannoni, and Patterson (2015), and Farhi and Werning (2019), the mortality risk gives rise to higher MPCs that are in line with empirical estimates.

[21]A natural restriction is that $\sum_g \pi_g \phi_g = 1$.

Without informational friction, the equilibrium outcomes are proportional to the fundamental. The magnitude of the responses depends on the MPCs and income exposures through via a Leontief inverse matrix in the consumption network.

Lemma 4.1. *In the economy without informational friction ($\tau_1 = \tau_2 = \infty$), the equilibrium outcomes are*

$$\begin{bmatrix} c_{1t}^* \\ c_{2t}^* \end{bmatrix} = \begin{bmatrix} c_1^* \\ c_2^* \end{bmatrix} r_t, \qquad \begin{bmatrix} c_1^* \\ c_2^* \end{bmatrix} = \left(\mathbf{I} - \begin{bmatrix} \frac{m_1\phi_1\pi_1}{1-(1-m_1)\rho} & \frac{m_1\phi_1\pi_2}{1-(1-m_1)\rho} \\ \frac{m_2\phi_2\pi_1}{1-(1-m_2)\rho} & \frac{m_2\phi_2\pi_2}{1-(1-m_2)\rho} \end{bmatrix} \right)^{-1} \begin{bmatrix} \frac{1-m_1}{1-(1-m_1)\rho} \\ \frac{1-m_2}{1-(1-m_2)\rho} \end{bmatrix},$$

and the aggregate output is

$$y_t^* = \sum_g \pi_g c_g^* r_t.$$

Now turn to the economy with incomplete information. The abundance in types of heterogeneity implies that the equilibrium outcomes will depend on higher-order expectations with rich structures. The following proposition summarizes the eventual dynamic pattern.

Proposition 4.3. *With heterogeneity in both information structure and MPCs, the aggregate output follows*

$$y_t = \left(\omega_1 \left(1 - \frac{\vartheta_1}{\rho}\right) \frac{1}{1-\vartheta_1 L} + \omega_2 \left(1 - \frac{\vartheta_2}{\rho}\right) \frac{1}{1-\vartheta_2 L} \right) y_t^*.$$

where ϑ_1 and ϑ_2 are the reciprocals of the outside roots of the determinant of $\mathbf{T}(L)$

$$\mathbf{T}(L) \equiv \begin{bmatrix} \frac{m_1\phi_1\pi_1}{1-(1-m_1)L^{-1}} & \frac{m_1\phi_1\pi_2}{1-(1-m_1)L^{-1}} \\ \frac{m_2\phi_2\pi_1}{1-(1-m_2)L^{-1}} & \frac{m_2\phi_2\pi_2}{1-(1-m_2)L^{-1}} \end{bmatrix} - \begin{bmatrix} \frac{(1-\rho L)(L-\rho)+\tau_1 L}{\tau_1 L} & 0 \\ 0 & \frac{(1-\rho L)(L-\rho)+\tau_2 L}{\tau_2 L} \end{bmatrix},$$

and ω_1 and ω_2 are constants that depend on deep parameters.

With incomplete information, both within-group and cross-group higher-order expectations about real interest rates matter for the output dynamics. Proposition 4.3 reveals that relative to the benchmark without informational friction, the aggregate outcome is now subject to a modification that depends on two AR(1) terms. These two persistence parameters $(\vartheta_1, \vartheta_2)$ capture the additional dynamics relative to the fundamental process of r_t. To determine $(\vartheta_1, \vartheta_2)$, it requires information from $\mathbf{T}(L)$: the GE consideration captured by the MPCs and income exposures interact with informational frictions captured by precision (τ_1, τ_2) when shaping the equilibrium dynamics. This result echos with our general characterization in section 3.4.

To gain further intuition on this interaction, we consider two special cases. In the first special case, we keep informational friction but assume away the heterogeneity in the friction, which helps highlight the role of heterogeneity in MPCs and income exposures.

Proposition 4.4. *With common information structure $\tau_1 = \tau_2 = \tau$, the aggregate output follows*

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

where ϑ is the reciprocal of the outside root of

$$T(L) = \sum_g \pi_g \frac{m_g \phi_g}{1 - (1 - m_g)L^{-1}} - \frac{(1 - \rho L)(L - \rho) + \tau L}{\tau L}.$$

1. *When $m_1 = m_2$, ϑ is independent of heterogeneous income exposures.*

2. *When $m_1 > m_2$, ϑ is increasing in the high MPC group's income exposure to aggregate output, ϕ_1.*

With common information structure, the effects of incomplete information can be represented by a single composite parameter ϑ, which is the case explored in Angeletos and Huo (2021). When consumers are with the same MPC, heterogeneity in income exposures to aggregate output is irrelevant: what matters is only the average common MPC.[22] When $m_1 > m_2$, increasing income exposure of the high MPC group to aggregate output amplifies the average dependence on aggregate output, which strengthens the general equilibrium consideration. This makes room for the higher-order expectations to play a more important role in shaping the outcome.

In the second special case, we maximize ϕ_1 by setting the low MPC group's income exposure to aggregate output to zero ($\phi_2 = 0$), while allowing heterogeneous information structures. This special case helps isolate the effects of heterogeneous information structure.

Proposition 4.5. *With $\phi_2 = 0$ and heterogeneous information structure, the group specific consumption follows*

$$c_{1t} - c_{1t}^* = -c_1^* \frac{\vartheta}{\rho} \frac{1}{1 - \vartheta L} \eta_t - m_1 \phi_1 \pi_2 \frac{1 - \frac{\vartheta}{\lambda_2}}{T(\lambda_2^{-1})(1 - (1 - m_1)\lambda_2)} \frac{1}{1 - \vartheta L} c_2^* \frac{\lambda_2}{\rho} \frac{1}{1 - \lambda_2 L} \eta_t \tag{4.4}$$

$$c_{2t} - c_{2t}^* = -c_2^* \frac{\lambda_2}{\rho} \frac{1}{1 - \lambda_2 L} \eta_t, \tag{4.5}$$

where ϑ is the reciprocal of the outside root of

$$T(L) \equiv \frac{m_1 \phi_1 \pi_1}{1 - (1 - m_1)L^{-1}} - \frac{(1 - \rho L)(L - \rho) + \tau_1 L}{\tau_1 L}.$$

To better understand this result, note that for the low MPC group, their consumption only depends on the first-order expectation about r_t. It follows that the deviation from the benchmark case only depends on the degree of informational friction (λ_2), and the GE consideration is irrelevant for group 2. In contrast, consumers in group 1 care about aggregate income. The within-group GE consideration is captured by the first term on the right-hand side of condition (4.4), and the cross-group GE

[22] Note that the economy is subject to the feasibility constraint $\sum_g \pi_g \phi_g = 1$.

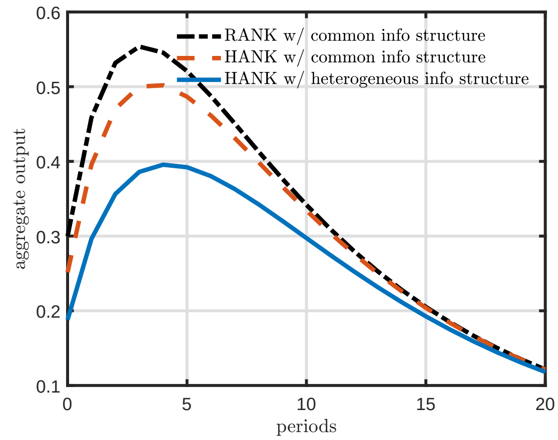


Figure 1: Responses of Output with Incomplete Information

Note: The persistence of the interest rate is set to be 0.9 and the two groups are with equal measure. In the RANK model (black broken line), $m_1 = m_2 = 0.3$ and the common precision is $\tau = 0.4$. In the HANK model with common information structure (red dashed line), $m_1 = 0.45 > m2 = 0.15$, $\phi_1 = 1.75 > \phi_2 = 0.25$, and $\tau_1 = \tau_2 = 0.4$. In the HANK model with heterogeneous information structure (blue solid line), $\tau_1 = 0.2 < \tau_2 = 0.6$ and the rest of the parameters are the same as that in the second case.

consideration is captured by the second term. Such additional GE considerations amplify the effects of information incompleteness. Therefore, when the high MPC group is subject to more informational friction than the low MPC group, it will have a larger quantitative bite on the aggregate output. These analytical results complement recent studies that explore the interaction between information heterogeneity and income heterogeneity in HANK models (Pfäuti and Seyrich, 2022; Guerreiro, 2022; Gallegos, 2023).

Finally, Figure 1 displays the impulse responses of output in three cases: (1) with common MPC and common information structure (black broken line), (2) with heterogeneous MPCs and common information structure (red dashed line), and (3) with heterogeneous MPCs and heterogeneous information structure (blue solid line). The responses are normalized by their counterparts without informational friction, which allows us to focus on the effects of incomplete information. Comparing case (1) and (2), with the same first-order expectation, heterogeneous MPCs and income exposures further dampens the impact response and induces additional sluggishness. Keeping the average first-order expectation the same as before, when the informational frictions is more severe for the high MPC group ($\tau_1 < \tau_2$), the effects of informational frictions on aggregate output are further amplified.

4.3 Reconciling with Empirical Evidence on Expectations

In this section, we show how our solution can be used to reconcile with empirical evidence on expectations, and how it differs from an alternative method in solving models with heterogeneous beliefs. To proceed, we adopt the model environment in Angeletos and La'O (2013) with decentralized trading and random matching. When shocks are persistent, Angeletos and La'O (2013) adopts a heterogeneous-prior approach to overcome the infinite regress problem. We instead maintain the rationality assumption.

In the economy, an individual agent i is endowed with a permanent fundamental κ_i, drawn from a normal distribution $\mathcal{N}(0,\sigma_\kappa^2)$. At the beginning of each period, an agent i is randomly matched with another agent indexed by $m(i,t)$. Agent i's optimal response is given by

$$a_{it} = \kappa_i + \alpha \mathbb{E}_{it}[a_{m(i,t)}], \tag{4.6}$$

where α captures the trade dependence between the two matched agents, and $a_{m(i,t)}$ is the action of agent i's match in period t.[23]

Besides knowing her own fundamental, agent i receives two signals every period about her trading partner

$$x_{it}^1 = \kappa_{m(i,t)} + \varepsilon_{it}, \tag{4.7}$$

$$x_{it}^2 = x_{m(i,t),t}^1 + \xi_t + u_{it}, \tag{4.8}$$

where $\varepsilon_{it} \sim \mathcal{N}(0,\sigma_\varepsilon^2)$ and $u_{it} \sim \mathcal{N}(0,\sigma_u^2)$ are both idiosyncratic noises, and ξ_t is a common noise. The fundamental of i's match in period t is $\kappa_{m(i,t)}$, which from i's perspective is also an i.i.d shock that follows $\mathcal{N}(0,\sigma_\kappa^2)$. As emphasized by Angeletos and La'O (2013), agent i's forecast about $\kappa_{m(i,t)}$ is pinned downed by i's first signal alone, and not affected by the second signal. However, agent i's forecast of $x_{m(i,t)t}^1$ and all the higher-order expectations are affected by the common noise ξ_t. The systematic variations in higher-order expectations induced by ξ_t generates fluctuations in aggregate outcomes. We assume that ξ_t follows a persistent process

$$\xi_t = \rho \xi_{t-1} + \eta_t. \tag{4.9}$$

Different from subsection 4.1, agent i has to form higher-order expectations about a random player $m(i,t)$ every period. Nevertheless, our method continues to work which yields the following equilibrium characterization.

Proposition 4.6. *The aggregate outcome a_t is given by*

$$a_t = \frac{\varphi}{1-\vartheta L}\eta_t \tag{4.10}$$

where

$$\vartheta = \frac{1}{2}\left[\frac{1}{\rho} + \rho + \frac{(1-\alpha)}{\rho(\sigma_\varepsilon^2+\sigma_u^2)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho(\sigma_\varepsilon^2+\sigma_u^2)}\right)^2 - 4}\,\right], \tag{4.11}$$

$$\varphi = \frac{\alpha^2\vartheta}{\rho}\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_u^2}\left(1-\alpha^2+\frac{\sigma_\varepsilon^2}{\sigma_\kappa^2}\left(1-\alpha^2\frac{\vartheta}{\rho}\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_u^2}\right)\right)^{-1}. \tag{4.12}$$

[23] See Angeletos and La'O (2013) for the details of the micro foundation.

Similar to Angeletos and La'O (2013), the common noise ξ_t leads to aggregate fluctuations, even though first-order expectations about the aggregate fundamental remain constant. In addition, when the common noise is persistent, the aggregate outcome inherits this property. Importantly, the persistence of former is necessarily greater than the latter which is determined in condition (4.11). This result also helps illustrate the working of Corollary 3.1: since ξ_t is not a fundamental in the best response, its persistence ρ will not enter the law of motion of the aggregate outcome.

Comparing with heterogeneous prior. A convenient device to avoid the infinite regress problem is to assume that agents have heterogeneous prior, as in Angeletos and La'O (2013) and Angeletos, Collard, and Dellas (2018). The heterogeneous prior assumption works as follows. Agent i observes both ξ_t and $a_{m(i,t)t}$ perfectly. However, agent i believes her match $m(i,t)$ observes a_i with a bias ξ_t. Supposing that agent i's policy rule is

$$a_{it} = f_1 a_i + f_2 a_{m(i,t)} + f_3 \xi_t,$$

then agent i believes that the action of her match is given by

$$\mathbb{E}_{it}[a_{m(i,t)}] = f_1 a_{m(i,t)} + f_2(a_i + \xi_t) + f_3 \xi_t.$$

By the method of undermined coefficients, it is straightforward to pin down the constants $\{f_1, f_2, f_3\}$ that satisfies the best response, which yields the following aggregate outcome with heterogeneous prior

$$a_t = \frac{\alpha^2}{(1-\alpha^2)(1-\alpha)} \xi_t. \tag{4.13}$$

The comparison between (4.10) and (4.13) echoes with our emphasis on the interaction between the GE consideration and informational friction. With heterogeneous priors, the aggregate outcome is perfectly correlated with ξ_t. This is in contrast with the result under rational expectations in which a different persistence by ϑ is endogenously determined. Meanwhile, the GE consideration α only modifies the impact response under heterogeneous priors, but it shapes the entire dynamics under rational expectations. The alternative approach with heterogeneous prior is convenient in obtaining the allocation, but at the cost of eliminating learning and higher-order expectations. Our approach helps reserve the role of dynamic higher-order expectations in shaping the aggregate outcome, without sacrificing the tractability.

Empirical evidence on expectations. The tractability also makes possible a clear mapping to the evidence on expectations. Consider the following regressions

$$\begin{aligned} a_{t+k} - \overline{\mathbb{E}}_t[a_{t+k}] &= K_{\mathrm{CG}}(\overline{\mathbb{E}}_t[a_{t+k}] - \overline{\mathbb{E}}_{t-1}[a_{t+k}]) + v_{t+k}, \\ a_{t+k} - \mathbb{E}_{it}[a_{t+k}] &= K_{\mathrm{BGMS}}(\mathbb{E}_{it}[a_{t+k}] - \mathbb{E}_{i,t-1}[a_{t+k}]) + v_{i,t+k}. \end{aligned}$$

The first regression proposed by Coibion and Gorodnichenko (2015) estimates the predictability of forecast errors using forecast revisions at the aggregate level, which detects deviations from rational expectations with common information if $K_{\text{CG}} \neq 0$. The second regression, proposed by Bordalo, Gennaioli, Ma, and Shleifer (2020), is at the individual level, and detects deviations from rationality if $K_{\text{BGMS}} \neq 0$.

Through the lens of our model, these two moments for one-period ahead forecast ($k = 1$) have the following properties.

Proposition 4.7. *1. With rational expectations, the coefficients K_{CG} and K_{BGMS} satisfy*

$$K_{CG} > K_{BGMS} = 0,$$

Furthermore, K_{CG} is decreasing in α.

2. With heterogeneous prior, the coefficients K_{CG} and K_{BGMS} are given by

$$K_{CG} = K_{BGMS} = \alpha - 1 < 0$$

Bordalo, Gennaioli, Ma, and Shleifer (2020) documents that $K_{\text{CG}} > 0$ and $K_{\text{CG}} > K_{\text{BGMS}}$.[24] This pattern is consistent with the model under rational expectations. With rational expectations, $K_{\text{BGMS}} = 0$ by construction. With dispersed information, an agent's forecast revision helps predict others' but not their own forecast error, which allows K_{CG} differs from zero. Furthermore, K_{CG} depends on the α. This suggests that in a GE setting where the outcome depends on the forecasts, one has to condition on the level of the GE consideration when mapping K_{CG} to the magnitude of the informational friction.

In contrast, with heterogeneous prior, all common shocks are publicly known. This implies that the moment at the individual level is always the same as that at the aggregate level. In addition, with the "naive" beliefs, agents always overreact to the news, implying a negative regression coefficient. The different implications on the properties of forecasts brings in additional caveats when substituting the rational-expectations framework with alternative approaches.

4.4 Integrating Belief Distortions with Dispersed Information

Recent work on expectations formation has also examined the assumption of individual rationality and provides evidence on significant deviation from this benchmark (Bordalo, Gennaioli, Ma, and Shleifer, 2020; Broer and Kohlhas, 2019). Different types of belief distortions have been proposed to account for the observed empirical patterns, including over/under confidence, over/under extrapolation, diagnostic expectations, and so on. Most of these studies focus on a partial equilibrium analysis, in the sense that the process of the variable to be forecast is taken to be exogenously given. This approach is effective in understanding how a certain type of belief distortion changes the properties

[24] Across different macroeconomic variables, about half of the K_{BGMS} coefficients are negative in Table 3 of Bordalo, Gennaioli, Ma, and Shleifer (2020) .

of individual or consensus forecasts, but it is not sufficient to evaluate its impact on the equilibrium outcomes. In this section, we show that our method can also be applied to models with bounded rationality, and help understand how forecasts with distorted beliefs and the endogenous outcomes are jointly determined in equilibrium.

To illustrate, we adopt the notion of diagnostic expectation formulation from BGMS. With rational expectations, the updating rule for the underlying state $\boldsymbol{z}_{it}$ in state space (3.7) is given by

$$\mathbb{E}_{it}[\boldsymbol{z}_{it}] = \mathbb{E}_{i,t-1}[\boldsymbol{z}_{it}] + \mathbf{K}\left(\boldsymbol{x}_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{z}_{it}]\right),$$

where $\mathbf{K}$ is the corresponding steady-state Kalman gain matrix. With diagnostic expectation, additional weight is put on the news, and the distorted belief denoted by $\widetilde{\mathbb{E}}[\cdot]$ is given by

$$\widetilde{\mathbb{E}}_{it}[\boldsymbol{z}_{it}] = \mathbb{E}_{i,t-1}[\boldsymbol{z}_{it}] + (1+\mu)\mathbf{K}\left(\boldsymbol{x}_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{z}_{it}]\right),$$

where $\mu \geq 0$ parameterizes the overreaction to the news. When $\mu = 0$, diagnostic expectations reduce to rational expectations.

Now consider the following beauty-contest model with diagnostic expectations

$$a_{it} = (1-\alpha)\widetilde{\mathbb{E}}_{it}[\xi_t] + \alpha\widetilde{\mathbb{E}}_{it}[a_t].$$

Following BGMS, we assume that the fundamental ξ_t follows an AR(1) process, and agents only observe a private signal $x_{it} = \xi_t + u_{it}$ every period where $u_{it} \sim \mathcal{N}(0, \sigma^2)$. Under this information structure, the diagnostic expectation about ξ_t is exactly the same as that in Proposition 1 of BGMS

$$\widetilde{\mathbb{E}}_{it}[\xi_t] = \mathbb{E}_{i,t-1}[\xi_t] + (1+\mu)\left(1 - \frac{\lambda}{\rho}\right)(x_{it} - \mathbb{E}_{i,t-1}[\xi_t]),$$

where λ is specified in equation (2.6). The new element here is the fixed point problem: the aggregate outcome a_t is endogenously determined by agents' diagnostic expectations, and agents have to form diagnostic expectations about this endogenous variable. The following proposition characterizes the fixed point in the equilibrium.

Proposition 4.8. *The aggregate outcome with diagnostic expectations is*

$$a_t = \frac{\lambda(1+\mu)}{\lambda(1+\mu) - \mu\vartheta}\left(1 - \frac{\mu}{1+\mu}\frac{\rho\vartheta}{\lambda}L\right)a_t^*, \tag{4.14}$$

where a_t^* is the outcome with rational expectations, and ϑ is a function of deep parameters

$$a_t^* = \left(1 - \frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\xi_t, \qquad \textit{and} \qquad \vartheta = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho\sigma^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho\sigma^2}\right)^2 - 4}\right].$$

Condition (4.14) gives the diagnostic-expectation dynamics as a transformation of the rational-expectation counterpart. This transformation consists of two parts: first, there is an additional response on impact captured by the constant term $\frac{\lambda(1+\mu)}{\lambda(1+\mu)-\mu\vartheta}$, which is due to the overreaction to the news. Note that the exact amount of overreaction also depends on the GE effects summarized in ϑ. Ceteris paribus, a stronger complementarity leads to a higher ϑ, and a more pronounced overreaction relative to its rational-expectation benchmark. Second, the diagnostic expectations also modify how to respond to past signals, which translates to modification of the entire dynamics, captured by the term $1 - \frac{\mu}{1+\mu}\frac{\rho\vartheta}{\lambda}L$. In this case, informational friction, GE feedback effect, and distorted belief jointly determine the outcome, and condition (4.14) neatly presents the role of each force.

The idea of combining incomplete information with certain kinds of distorted beliefs in a general equilibrium setting goes beyond the example presented above. For example, Angeletos, Huo, and Sastry (2021) combines over-extrapolation and over-/under-confidence with dispersed information to account for the identified delayed overshooting of consensus forecast in response to business-cycle shocks. We expect our method helps to facilitate more interaction between bounded rationality and incomplete information.

5. Endogenous Information

So far, we have maintained the assumption that the signal process is exogenously determined and independent of agents' actions. In this section we consider the case where signals contain variables that are endogenously determined in equilibrium. We first discuss how the models with endogenous information are related to those with exogenous information. We then discuss specific examples where the finite-state representation no longer exists. Finally, we propose a numerical algorithm to compute models with endogenous information.

In contrast with the previous setup, we modify the signal structure in the following way

$$\boldsymbol{x}_{it} = \mathbf{M}(L)\boldsymbol{s}_{it} + \mathbf{P}(L)\boldsymbol{a}_t. \tag{5.1}$$

The new element $\mathbf{P}(L)\boldsymbol{a}_t$ allows the signal to depend on the aggregate outcomes, and therefore the informativeness of the signal is endogenously determined in equilibrium. For example, agents could learn the aggregate state by observing past prices, outputs, and so on, which are also outcomes of agents' decisions. We define the equilibrium as follows.

Definition 2. *A linear Bayesian-Nash equilibrium with endogenous information is a policy rule $\boldsymbol{h}(L)$ for agents and a law of motion $\boldsymbol{\mathcal{H}}(L)$ for the aggregate outcome, such that*

1. *The individual action $\boldsymbol{a}_{it} = \boldsymbol{h}(L)\boldsymbol{x}_{it}$ satisfies the best response (3.1), taking the following exogenous signal process as given*
$$\boldsymbol{x}_{it} = \mathbf{M}(L)\boldsymbol{s}_{it} + \mathbf{P}(L)\boldsymbol{\mathcal{H}}(L)\boldsymbol{\Lambda}\boldsymbol{s}_{it}. \tag{5.2}$$

The aggregate outcome is consistent with individual actions: $\boldsymbol{a}_t = \int \boldsymbol{a}_{it}$.

2. *The aggregate outcome is consistent with the law of motion of the signal*

$$\boldsymbol{a}_t = \boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\boldsymbol{s}_{it} = \boldsymbol{\mathcal{H}}(L)\boldsymbol{\Lambda}\boldsymbol{s}_{it}.$$

Note that there are two distinct consistency requirements: the perceived law of motion $\boldsymbol{\mathcal{H}}(L)$ has to be the same as that enters the best response and the signal process. Accordingly, we purposely separate the equilibrium definition into two parts: in part (1), given a particular perceived law of motion $\boldsymbol{\mathcal{H}}(L)$, agents solve for their optimal policy $\boldsymbol{h}(L)$ in the same way as in an exogenous-information economy, and the results from Section 3 can be applied. In part (2), the additional consistency requirement on the signal process is unique to the endogenous-information equilibrium.

This definition also makes it clear that the endogenous-information equilibrium is a *particular* exogenous information equilibrium. From individual agents' perspectives, the competitive nature of the equilibrium implies that the information process is always exogenous to them. However, this argument does not mean that the distinction between exogenous and endogenous information is irrelevant, as only in the latter case the informativeness of signals varies with changes in policies, technologies, and market structures.

5.1 Infinite-State Representation

Notably, with endogenous information, the equilibrium may not admit a finite-state representation. In this subsection, we provide such an example which is a natural extension of the models with exogenous information.

Example. Suppose the best response is the same as that in Section 2

$$a_{it} = (1-\alpha)\mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[a_t],$$

where $\xi_t = \rho\xi_{t-1} + \eta_t$. Different from previous cases, agents receive an exogenous signal and an endogenous signal with i.i.d private noises every period

$$x^1_{it} = \eta_t + u_{it}, \quad \text{and} \quad x^2_{it} = a_t + \varepsilon_{it}.$$

The inclusion of the aggregate outcome a_t makes the second signal endogenous. There are three shocks and two signals, and the signal system is non-square.

Despite the seemingly insignificant deviation from previous examples by including the aggregate outcome in the signal process, the equilibrium can become much more complex to characterize, as shown in the following proposition.

Proposition 5.1. *Assume $\alpha \in (-1, 1)$ and all the shocks are with positive variances. Then, in the above example, the aggregate outcome in equilibrium never admits a finite-state representation.*

The basic idea behind the infinite-state result can be understood via Corollary 3.1. With endogenous information, the equilibrium process $a_t = \mathcal{H}(L)\eta_t$ itself enters the signal process $\mathbf{M}(L)$. At the same time, the coefficients of $\mathcal{H}(L)$ need to be consistent with the determinant of the corresponding $\mathbf{T}(L)$ in equation (3.14)—a function of $\mathbf{M}(L)$. However, such consistence in the fixed-point problem cannot be achieved with a finite-state representation. As long as the guess of the equilibrium is restricted to be within the realm of finite ARMA processes, Proposition 5.1 formally establishes that any such conjecture with a finite-state process cannot be supported as an equilibrium.[25]

It is important to note that the infinite-state result is not due to higher-order expectations per se. If the perceived law of motion for a_t follows a finite-order process, Theorem 1 implies the actual law of motion also follows a finite-order process, despite the dependence on higher-order expectations. With endogenous information, the additional complication lies in that the signal process itself cannot be represented as a finite-order process.

Proposition 5.1 also implies that such infinite-state result does not hinge on explicit coordination motive. Even when $\alpha = 0$, agents still need to forecast the action of others as it appears in their signal. The endogenous signal therefore introduces an implicit form of coordination, which leads to the type of fixed-point problem in Definition 2.

In the literature, a forward-looking asset pricing example is provided by Makarov and Rytchkov (2012). Their proof is based on the orthogonality condition of projection obtained from the inverse z-transform, which leads to the necessary conditions to admit Markovian dynamics. We adopt a different proof strategy, which relies on the properties of the fundamental representation, the Wiener-Hopf prediction formula, and the annihilation operator. These properties can be used to study more complex problems. For example, in Appendix A.20, we have shown that the infinite-state result extends to the case in which the fundamental follows an arbitrary AR (p) process.[26]

5.2 Finite-State Approximation

In this subsection, we provide an algorithm that approximates the aggregate outcome with a relatively low-order ARMA process. The algorithm is based on the equilibrium Definition 2. The key idea is to utilize our results in Section 3 to solve the exogenous-information equilibrium, which helps save the required state variables.

[25] With this conjecture, one may think that keep track of $\{x^1_{it} = \eta_t + u_{it}, x^2_{it} - \rho x^2_{i,t-1} = \delta\eta_t + \varepsilon_{it} - \rho\varepsilon_{i,t-1}\}$ every period is sufficient to make the inference about η_t as $\eta_t, \varepsilon_{it}, \varepsilon_{i,t-1}$ are i.i.d shocks. This argument fails to recognize that the forecast can be improved by using additional signals from the past. For example, since $x^1_{i,t-1}$ is helpful in inferring η_{t-1}, it is also helpful in inferring ξ_{t-1} and $\varepsilon_{i,t-1}$. As a result, $x^1_{i,t-1}$ should be used in forecasting η_t. By the same logic, all past signals are relevant.

[26] We are grateful to Kyle Jurado for pointing out a gap in an early version of the proof.

To illustrate the strategy, consider on the following framework

$$a_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[a_t] + \gamma \mathbb{E}_{it}[a_{t+1}] + \beta \mathbb{E}_{it}[a_{i,t+1}], \tag{5.3}$$

where $\xi_t = \rho \xi_{t-1} + \eta_t$, and the signal process is given by

$$x_{it}^1 = \xi_t + u_{it}, \quad \text{and} \quad x_{it}^2 = a_t + \varepsilon_{it}.$$

The best response (5.3) allows for both static and forward-looking complementarity. This structure nests a number of commonly used models in the literature, and the algorithm can be applied for more general best responses as well.[27]

Numerical Algorithm. When information is endogenous, the finite-state representation no longer holds in general. We proceed with an iterative algorithm that maps from the perceived signal process to the actual law of motion.

Starting with a perceived process $a_t = \mathcal{H}^{(0)}(L)\xi_t$ that admits an ARMA representation, we compute a particular exogenous-information equilibrium where the *exogenous* signals are given by

$$x_{it}^1 = \xi_t + u_{it}, \quad \text{and} \quad x_{it}^2 = \mathcal{H}^{(0)}(L)\xi_t + \varepsilon_{it}.$$

Denote the law of motion in the exogenous-information equilibrium as $a_t = \mathcal{H}^{(1)}(L)\xi_t$, which can be obtained from our earlier results in Section 3. Importantly, for individual agents, their perceived aggregate outcome that enters the best response is $\mathcal{H}^{(1)}(L)\xi_t$, rather than $\mathcal{H}^{(0)}(L)\xi_t$ that enters the signal process.[28]

Though by construction, the perceived actual law of motion $\mathcal{H}^{(1)}(L)\xi_t$ is consistent with the actual law of motion, it may not satisfy part (2) in Definition 2. One can therefore set the next perceived law of motion for the signal process to be $\mathcal{H}^{(1)}(L)$, and iterate this process until $||\mathcal{H}^{(k+1)}(L) - \mathcal{H}^{(k)}(L)||$ is small enough.

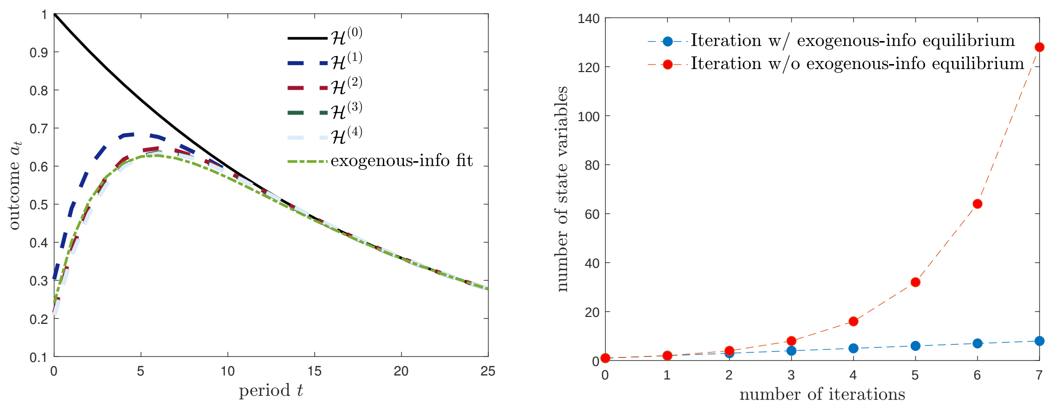
We illustrate two advantages of this numerical algorithm: fast convergence speed and efficiency in saving state variables. As a comparison, consider the following alternative iteration algorithm without solving an exogenous-information equilibrium in each iteration:

$$a_t = \overline{\mathbb{E}}_t[\xi_t] + \alpha \overline{\mathbb{E}}_t \left[\mathcal{H}^{(0)}(L)\xi_t\right] + \gamma \overline{\mathbb{E}}_t \left[\mathcal{H}^{(0)}(L)\xi_{t+1}\right] + \beta \mathbb{E}_{it}[a_{i,t+1}]. \tag{5.4}$$

Different from our original mapping, $\mathcal{H}^{(0)}(L)$ enters both the signal process and the best response.

[27] For example, by setting $\gamma = \beta = 0$, it nests the static beauty contests in Morris and Shin (2002), Woodford (2003), Maćkowiak and Wiederholt (2009) and Angeletos and La'O (2010); by allowing $\beta > 0$ and $\gamma > 0$, it nests the forward-looking beauty contests in Allen, Morris, and Shin (2006), Nimark (2008), Nimark (2017), and Angeletos and Huo (2021). The method also works for environments with backward-looking best responses.

[28] This distinction is irrelvant in the true equilibrium where $\mathcal{H}^{(0)}(L) = \mathcal{H}^{(1)}(L)$. In the iterative algorithm, such distinction helps speed up the convergence.



(a) Convergence of IRF

(b) Required number of state variables

Figure 2: Numerical Properties in the Endogenous Information Example

Parameters: $\rho = 0.95, \alpha = 0.5, \gamma = 0.2, \beta = 0.1, \sigma_\eta = 1, \sigma_u = \sigma_\varepsilon = 2.$

Consequently, there is no need to solve for the exogenous-information equilibrium, and the actual law of motion $a_t = \mathcal{H}^{(1)}(L)\eta_t$ is different from agents' perception. The left panel of Figure 2 displays the convergence paths of the impulse response functions when solving the exogenous-information equilibrium in the iteration. Starting with the perfect information solution, the law of motion converges after a small number of iterations. From the second iteration, the IRF of aggregate outcomes can hardly be distinguished from its further iterations. The right panel of Figure 2 compares the required number of state variables with the aforementioned alternative algorithm. The number of required minimal state variables increases linearly in our approach, but increases exponentially in this alternative method.

This numerical algorithm complements the literature on computing models with endogenous dispersed information. To deal with the issue of potential infinite history, a common strategy is to truncate the history as in Hellwig (2002), Lorenzoni (2009), and Venkateswaran (2014). Nimark (2017) instead approximates the equilibrium outcome with a finite-order expectations of the fundamental, and this method has an interesting bounded rationality interpretation. Both of these methods are flexible and straightforward to implement, but typically require a relatively large state space to for an accurate approximation. Recently, Han, Tan, and Wu (2019) build on the bridge between Kalman filter and Wiener filter proposed in Section 3 to obtain the fundamental representation, and they implement the inference with the discrete Fourier transform which helps speed up the annihilating operator. These different methods have their own comparative advantage, and researchers could adopt the best suited one for their particular application.

6. Conclusion

We develop a method that helps to solve and characterize the equilibrium outcomes when information is incomplete and agents coordinate with each other. The key step is to combine the Kalman filter and Wiener filter to make the inference problem traceable. We show that the equilibrium outcome always admits a finite-state representation when signals follow finite ARMA processes, and we characterize how the endogenous persistence depends on the interaction between information incompleteness and general equilibrium consideration. We also illustrate how to use the method to compute the equilibrium with endogenous information.

We demonstrate that our method can help derive applied lessons in a sequence of applications. Particularly, it is flexible enough to accommodate deviations from strong rationality, such as diagnostic expectations (Bordalo, Gennaioli, Ma, and Shleifer, 2020), over-/under-confidence (Broer and Kohlhas, 2019), and over-/under-extrapolation (Greenwood and Shleifer, 2014). These belief distortions have been shown to be necessary to rationalize the evidence on expectations. The method allows one to explore their general equilibrium implications under incomplete information.

Another direction of future research includes the exploration of how the structure of dynamic coordination affects the equilibrium outcome. The applications in this paper have focused on static and Euler-type forward-looking complementarities, but the method allows one to consider much more sophisticated dependence on past and future aggregate outcomes estimated from micro data. Our results can then assist in building the bridge from micro to macro.

References

Acharya, S. (2013): "Dispersed Beliefs and Aggregate Demand Management," Working paper.

Acharya, S., J. Benhabib, and Z. Huo (2021): "The anatomy of sentiment-driven fluctuations," Forthcoming, *Journal of Economic Theory*.

Allen, F., S. Morris, and H. S. Shin (2006): "Beauty contests and iterated expectations in asset markets," *The Review of Financial Studies*, 19(3), 719–752.

Angeletos, G.-M., F. Collard, and H. Dellas (2018): "Quantifying Confidence," *Econometrica*, 86(5), 1689–1726.

Angeletos, G.-M., and Z. Huo (2021): "Myopia and Anchoring," *American Economic Review*, 111(4), 1166–1200.

Angeletos, G.-M., Z. Huo, and K. Sastry (2021): "Imperfect Macroeconomic Expectations: Evidence and Theory," in *NBER Macroeconomics Annual 2020*, vol. 35, chap. 1.

Angeletos, G.-M., and J. La'O (2010): "Noisy Business Cycles," in *NBER Macroeconomics Annual 2009, Volume 24*, pp. 319–378. University of Chicago Press.

——— (2013): "Sentiments," *Econometrica*, 81(2), 739–779.

Angeletos, G.-M., and C. Lian (2016): "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," *Handbook of Macroeconomics*, 2, 1065–1240.

Angeletos, G.-M., and C. Lian (2018): "Forward guidance without common knowledge," *American Economic Review*, 108(9), 2477–2512.

Angeletos, G.-M., and A. Pavan (2007): "Efficient Use of Information and Social Value of Information," *Econometrica*, 75(4), 1103–1142.

Auclert, A., M. Rognlie, and L. Straub (2020): "Micro Jumps, Macro Humps: monetary policy and business cycles in an estimated HANK model," *mimeo*.

Bacchetta, P., and E. V. Wincoop (2006): "Can Information Heterogeneity Explain the Exchange Rate Determination Puzzle?," *American Economic Review*, 96(3), 552–576.

Barsky, R. B., and E. R. Sims (2012): "Information, Animal Spirits, and the Meaning of Innovations in Consumer Confidence," *American Economic Review*, 102(4), 1343–77.

Benhabib, J., P. Wang, and Y. Wen (2015): "Sentiments and Aggregate Demand Fluctuations," *Econometrica*, 83, 549–585.

Bergemann, D., T. Heumann, and S. Morris (2017): "Information and Interaction," Discussion paper, Yale University, Cowles Foundation Discussion Paper No. 2088.

Bergemann, D., and S. Morris (2013): "Robust Predictions in Games With Incomplete Information," *Econometrica*, 81(4), 1251–1308.

Bianchi, F., C. L. Ilut, and H. Saijo (2021): "Implications of Diagnostic Expectations: Theory and Applications," Discussion paper, National Bureau of Economic Research.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer (2020): "Overreaction in Macroeconomic Expectations," *American Economic Review*, 110(9), 2748–82.

Bordalo, P., N. Gennaioli, A. Shleifer, and S. J. Terry (2021): "Real credit cycles," Discussion paper, National Bureau of Economic Research.

Broer, T., and A. Kohlhas (2019): "Forecaster (Mis-)Behavior," *IIES miméo*.

Broer, T., A. Kohlhas, K. Mitman, and K. Schlafmann (2021): "Information and wealth heterogeneity in the macroeconomy," .

Chahrour, R., and G. Gaballo (2016): "Learning from Prices: Amplification and Business Fluctuations," Discussion paper.

Chahrour, R., and K. Jurado (2018): "News or noise? The missing link," *American Economic Review*, 108(7), 1702–36.

Chahrour, R., K. Nimark, and S. Pitschner (2021): "Sectoral media focus and aggregate fluctuations," *American Economic Review*, 111(12), 3872–3922.

Chiang, Y.-T. (2022): "Attention and Fluctuations in Macroeconomic Uncertainty," Discussion Paper 2022-4, FRB St. Louis Working Paper.

Coibion, O., and Y. Gorodnichenko (2015): "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts," *American Economic Review*, 105(8), 2644–78.

Fuhrer, J. C. (2018): "Intrinsic expectations persistence: evidence from professional and household survey expectations," Working Paper.

Gallegos, J.-E. (2023): "HANK Beyond FIRE: Amplification, Forward Guidance and Belief Shocks," .

Golub, B., and S. Morris (2019): "Expectations, Networks and Conventions," *mimeo*.

Graham, L., and S. Wright (2010): "Information, Heterogeneity and Market Incompleteness," *Journal of Monetary Economics*, 57(2), 164–174.

Greenwood, R., and A. Shleifer (2014): "Expectations of returns and expected returns," *The Review of Financial Studies*, 27(3), 714–746.

Guerreiro, J. (2022): "Belief Disagreement and Business Cycles," .

Hamilton, J. D. (1994): *Time Series Analysis*. Princeton University Press, Princeton, 2 edn.

Han, Z., F. Tan, and J. Wu (2019): "Analytic Policy Function Iteration," working paper.

Hellwig, C. (2002): "Public announcements, adjustment delays and the business cycle," mimeo, UCLA.

Hellwig, C., and V. Venkateswaran (2009): "Setting the right prices for the wrong reasons," *Journal of Monetary Economics*, 56, S57–S77.

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

Huo, Z., and N. Takayama (2022): "Higher-Order Beliefs, Confidence and Business Cycles," Manuscript, Yale University.

Kasa, K. (2000): "Forecasting the Forecasts of Others in the Frequency Domain," *Review of Economic Dynamics*, 3(4), 726–756.

Kasa, K., T. B. Walker, and C. H. Whiteman (2014): "Heterogeneous Beliefs and Tests of Present Value Models," *Review of Economic Studies*, 81(3), 1137–1163.

Kohlhas, A., and A. Walther (2021): "Asymmetric attention," Forthcoming, *American Economic Review*.

La'O, J., and A. Tahbaz-Salehi (2020): "Optimal Monetary Policy in Production Networks," *mimeo*.

L'Huillier, J.-P., S. R. Singh, and D. Yoo (2021): "Diagnostic Expectations and Macroeconomic Volatility," mimeo.

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

Lucas, R. E. (1972): "Expectations and the Neutrality of Money," *Journal of economic theory*, 4(2), 103–124.

Maćkowiak, B., and M. Wiederholt (2009): "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 99(3).

——— (2015): "Business Cycle Dynamics under Rational Inattention," *Review of Economic Studies*, 82(4), 1502–1532.

Makarov, I., and O. Rytchkov (2012): "Forecasting the forecasts of others: Implications for asset pricing," *Journal of Economic Theory*, 147(3), 941–966.

Mankiw, G. N., and R. Reis (2002): "Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 1295–6.

Melosi, L. (2014): "Estimating models with dispersed information," *American Economic Journal: Macroeconomics*, 6(1), 1–31.

——— (2016): "Signalling effects of monetary policy," *The Review of Economic Studies*, 84(2), 853–884.

Miao, J., J. Wu, and E. R. Young (2021): "Macro-financial volatility under dispersed information," *Theoretical Economics*, 16(1), 275–315.

Morris, S., and H. S. Shin (2002): "Social Value of Public Information," *American Economic Review*, 92(5), 1521–1534.

Nimark, K. (2008): "Monetary policy with signal extraction from the bond market," *Journal of Monetary Economics*, 55(8), 1389–1400.

——— (2017): "Dynamic higher order expectations," Working paper.

Nimark, K. P. (2014): "Man-bites-dog business cycles," *American Economic Review*, 104(8), 2320–67.

Onatski, A. (2006): "Winding number criterion for existence and uniqueness of equilibrium in linear rational expectations models," *Journal of Economic Dynamics and Control*, 30(2), 323–345.

Patterson, C. (2019): "The Matching Multiplier and the Amplification of Recessions," *Northwestern University mimeo*.

Pfäuti, O., and F. Seyrich (2022): "A behavioral heterogeneous agent new keynesian model," .

Rondina, G. (2008): "Incomplete Information and Informative Pricing: Theory and Application," mimeo.

Rondina, G., and T. B. Walker (2021): "Confounding Dynamics," Forthcoming, *Journal of Economic Theory*.

Rozanov, Y. A. (1967): *Stationary random processes*. Holden-Day, San Francisco.

Rozsypal, F., and K. Schlafmann (2022): "Overpersistence Bias in Individual Income Expectations and its Aggregate Implications," *American Economic Journal: Macroeconomics (Forthcoming)*.

Sargent, T., and L. Hansen (1981): *Linear Rational Expectations Models of Dynamically Interrelated Variables*vol. 1, pp. 127–56. University of Minnesota Press.

Sargent, T. J. (1991): "Equilibrium with signal extraction from endogenous variables," *Journal of Economic Dynamics and Control*, 15(2), 245–273.

Sims, C. A. (2007): "On the genericity of the winding number criterion for linear rational expectations models," Discussion paper.

Tan, F., and T. B. Walker (2015): "Solving generalized multivariate linear rational expectations models," *Journal of Economic Dynamics and Control*, 60, 95–111.

Taub, B. (1989): "Aggregate fluctuations as an information transmission mechanism," *Journal of Economic Dynamics and Control*, 13(1), 113–150.

Townsend, R. M. (1983): "Forecasting the Forecasts of Others," *Journal of Political Economy*, 91(4), 546–88.

Venkateswaran, V. (2014): "Heterogeneous Information and Labor Market Fluctuations," Working paper.

Weinstein, J., and M. Yildiz (2007): "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 75(2), 365–400.

Whiteman, C. H. (1983): *Linear Rational Expectations Models: A User's Guide*. University of Minnesota Press.

Whittle, P. (1963): *Prediction and regulation by linear least-square methods*. English Univ. Press.

Woodford, M. (2003): "Imperfect Common Knowledge and the Effects of Monetary Policy," *Knowledge, Information, and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, p. 25.

Online Appendix

A. Proof of Theorems and Propositions

A.1 Proof of Lemma 3.1

Proof. The proof follows directly from Chapter 13.1 in Hamilton (1994). □

A.2 Proof of Wiener-Hopf Prediction Formula

Proof. A formal proof can be found in Whittle (1963). □

A.3 Formula for Annihilation Operator

To handle the annihilation operator, we prove the following lemma.

Lemma A.1. *Suppose $g(L)$ does not contain negative powers of L in expansion and $|\xi_k| < 1$.*

$$\left[\frac{g(L)}{\prod_{k=1}^{\ell}(L-\xi_k)^{a_k}}\right]_+ = \frac{g(L)}{\prod_{k=1}^{\ell}(L-\xi_k)^{a_k}} - \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(z-\xi_k)^{a_k-\tau}}\left[\frac{g(L)}{\prod_{h\neq k}(L-\xi_h)^{a_h}}\right]^{(\tau)}_{L=\xi_k}. \tag{A.1}$$

Proof. We are able to write

$$\frac{g(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}} = \tilde{g}(z) + \frac{\tilde{f}(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}}$$

where $\tilde{g}(z)$ and $\tilde{f}(z)$ are polynomial, ζ_i are distinct and $|\zeta_i| < 1$, and we can choose the order of $\tilde{f}(z)$ to be strictly less than the order of $\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}$. It follows that the partial fraction decomposition of the RHS can be written as

$$\tilde{g}(z) + \frac{\tilde{f}(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}} = \tilde{g}(z) + \sum_{k=1}^{\ell}\sum_{\tau=1}^{a_k}\frac{c_{k,\tau}}{(z-\xi_k)^{\tau}} + \sum_{i=1}^{\nu}\sum_{\tau=1}^{b_i}\frac{d_{i,\tau}}{(1-\zeta_i z)^{\tau}} \tag{A.2}$$

where $c_{k,\tau}$ and $d_{i,\tau}$ are unknown real constants to be determined. Then,

$$\left[\frac{g(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}}\right]_+ = \left[\tilde{g}(z) + \sum_{k=1}^{\ell}\sum_{\tau=1}^{a_k}\frac{c_{k,\tau}}{(z-\xi_k)^{\tau}} + \sum_{i=1}^{\nu}\sum_{\tau=1}^{b_i}\frac{d_{i,\tau}}{(1-\zeta_i z)^{\tau}}\right]_+ = \frac{g(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}} - \sum_{k=1}^{\ell}\sum_{\tau=1}^{a_k}\frac{c_{k,\tau}}{(z-\xi_k)^{\tau}}.$$

Thus, what we ultimately need to obtain is $c_{h,\tau}$ where $h \in \{1,\ldots,\ell\}$ and $\tau \in \{1,\ldots,a_h\}$. Let us multiply $(z-\xi_h)^{a_h}$ both sides of (A.2):

$$(z-\xi_h)^{a_h}\tilde{g}(z) + \frac{(z-\xi_h)^{a_h}\tilde{f}(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}} = (z-\xi_h)^{a_h}\tilde{g}(z) + \sum_{k=1}^{\ell}\sum_{\tau=1}^{a_k}\frac{(z-\xi_h)^{a_h}c_{k,\tau}}{(z-\xi_k)^{\tau}} + \sum_{i=1}^{\nu}\sum_{\tau=1}^{b_i}\frac{(z-\xi_h)^{a_h}d_{i,\tau}}{(1-\zeta_i z)^{\tau}},$$

and define

$$f_h(z) \quad \equiv \quad (z-\xi_h)^{a_h}\tilde{g}(z) + \frac{(z-\xi_h)^{a_h}\tilde{f}(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}} = \frac{g(z)}{\prod_{k\neq h}(z-\xi_k)^{a_k}}.$$

Now, we invoke the Heaviside expansion theorem and by i-times ($i \in \{0, 1, ..., a_h - 1\}$) differentiating both side with respect to z and evaluating at $z = \xi_h$, the LHS and RHS of (A.3) are

$$\begin{aligned}
\text{(LHS)} \quad &= \quad \frac{d^i}{dz^i}\left((z-\xi_h)^{a_h}\tilde{g}(z) + \frac{(z-\xi_h)^{a_h}\tilde{f}(z)}{\prod_{k=1}^{\ell}(z-\xi_k)^{a_k}\prod_{i=0}^{\nu}(1-\zeta_i z)^{b_i}}\right)\Bigg|_{z=\xi_h} = f_h^{(i)}(\xi_h),\\
\text{(RHS)} \quad &= \quad \frac{d^i}{dz^i}\left((z-\xi_h)^{a_h}\tilde{g}(z) + \sum_{k=1}^{\ell}\sum_{\tau=1}^{a_k}\frac{(z-\xi_h)^{a_h}c_{k,\tau}}{(z-\xi_k)^{\tau}} + \sum_{i=1}^{\nu}\sum_{\tau=1}^{b_i}\frac{(z-\xi_h)^{a_h}d_{i,\tau}}{(1-\zeta_i z)^{\tau}}\right)\Bigg|_{z=\xi_h}\\
&= \quad 0 + \frac{d^i}{dz^i}\left((z-\xi_h)^{a_h-1}c_{h,1} + \ldots + c_{h,a_h}\right)\Big|_{z=\xi_h} + 0\\
&= \quad \frac{(a_h-1)!(z-\xi_h)^{a_h-i-1}c_{h,1}}{(a_h-i-1)!} + \ldots + \frac{(i+1)!(z-\xi_h)c_{h,a_h-i-1}}{1!} + i!c_{h,a_h-i}\Bigg|_{z=\xi_h} = i!c_{h,a_h-i},
\end{aligned}$$

respectively and thus c_{h,a_h} is determined as $c_{h,a_h-i} = \frac{f_h^{(i)}(\xi_h)}{i!}$. □

A.4 Proof of Proposition 3.1

From the Fundamental Representation Theorem, we have that

$$\mathbf{B}(L)^{-1} = \mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L,$$

and from Lemma 3.1 it follows that

$$\boldsymbol{z}_{it} = (\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi}\boldsymbol{s}_t, \quad \text{and,} \quad \mathbf{x}_{it} = \mathbf{H}\boldsymbol{z}_{it} + \boldsymbol{\Psi}\boldsymbol{s}_{it},$$

which implies

$$\mathbf{x}_{it} = \mathbf{M}(L)\boldsymbol{s}_{it} = (\mathbf{H}\,(\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi} + \boldsymbol{\Psi})\boldsymbol{s}_{it}.$$

Hence,

$$\begin{aligned}
\mathbf{B}(L)^{-1}\mathbf{M}(L) &= \left(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L\right)\left(\mathbf{H}\,(\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi} + \boldsymbol{\Psi}\right)\\
&= \mathbf{H}\left(\mathbf{I} - \left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FKH}L\right)(\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi} + \left(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L\right)\boldsymbol{\Psi}\\
&= \mathbf{H}\left(\mathbf{I} - \left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\left((\mathbf{I} - \mathbf{F}L) + \mathbf{FKH}L - (\mathbf{I} - \mathbf{F}L)\right)\right)(\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi} + \left(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L\right)\boldsymbol{\Psi}\\
&= \mathbf{H}\left(\mathbf{I} - \mathbf{I} + \left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}(\mathbf{I} - \mathbf{F}L)\right)(\mathbf{I} - \mathbf{F}L)^{-1}\,\boldsymbol{\Phi} + \left(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L\right)\boldsymbol{\Psi}\\
&= \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\boldsymbol{\Phi} + \left(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right]^{-1}\mathbf{FK}L\right)\boldsymbol{\Psi}\\
&= \frac{\mathbf{H}\,\mathrm{Adj}\,(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L)\,(\boldsymbol{\Phi} - \mathbf{FK}\boldsymbol{\Psi}L) + \boldsymbol{\Psi}\det(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L)}{\det(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L)}
\end{aligned}$$

$$= \frac{\mathbf{H}\,\mathrm{Adj}\left(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right)(\boldsymbol{\Phi} - \mathbf{FK}\boldsymbol{\Psi}L) + \boldsymbol{\Psi}\prod_{k=1}^{u}(1-\lambda_k L)}{\prod_{k=1}^{u}(1-\lambda_k L)},$$

where $\{\lambda_k\}_{k=1}^{u}$ are the non-zero eigenvalues of $\mathbf{F} - \mathbf{FKH}$. The last equality follows from the fact that

$$\det\left(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L\right) = L^{v}\det\left(\mathbf{I}L^{-1} - (\mathbf{F} - \mathbf{FKH})\right) = L^{v}L^{-(v-u)}\prod_{k=1}^{u}\left(L^{-1} - \lambda_k\right) = \prod_{k=1}^{u}(1-\lambda_k L),$$

where v is the dimension of $\mathbf{F}$. It follows that,

$$\begin{aligned}
\mathbf{B}(L^{-1})^{-1}\mathbf{M}\left(L^{-1}\right) &= \frac{\mathbf{H}\,\mathrm{Adj}\left(\mathbf{I} - (\mathbf{F} - \mathbf{FKH})\,L^{-1}\right)(\boldsymbol{\Phi} - \mathbf{FK}\boldsymbol{\Psi}L^{-1}) + \boldsymbol{\Psi}\prod_{k=1}^{u}\left(1-\lambda_k L^{-1}\right)}{\prod_{k=1}^{u}(1-\lambda_k L^{-1})} \\
&= \frac{\mathbf{H}\,\mathrm{Adj}\left(\mathbf{I}L - (\mathbf{F} - \mathbf{FKH})\right)(\boldsymbol{\Phi}L - \mathbf{FK}\boldsymbol{\Psi}) + L^{v-u}\boldsymbol{\Psi}\prod_{k=1}^{u}(L-\lambda_k)}{L^{v-u}\prod_{k=1}^{u}(L-\lambda_k)} \\
&\equiv \frac{\mathbf{G}'(L)}{L^{v-u}\prod_{k=1}^{u}(L-\lambda_k)}.
\end{aligned}$$

With Lemma A.1, it is straightforward to show that the optimal prediction is given by

$$\begin{aligned}
&\left[\phi(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_{+}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it} = \phi(L)\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\boldsymbol{x}_{it} \\
&- \left(\sum_{k=1}^{u}\frac{\phi(\lambda_k)\mathbf{G}(\lambda_k)}{(L-\lambda_k)\lambda_k^{v-u}\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)} + \sum_{k=0}^{v-u-1}\frac{1}{k!L^{v-u-k}}\left[\frac{\phi(L)\mathbf{G}(L)}{\prod_{\tau=1}^{u}(L-\lambda_\tau)}\right]_{L=0}^{(k)}\right)\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it}.
\end{aligned}$$

A.5 Proof of Lemma 3.2

We proceed in two steps. We fist construct the matrix $\mathbf{T}(L)$ using the Wiener-Hopf prediction formula, and then we determine the number of endogenous constants that need to be determined.

Part I: constructing $\mathbf{T}(L)$. Following Assumption 1, we specify the relevant inside poles in the following way

$$\begin{aligned}
(\boldsymbol{\beta}(L) - \mathbf{I}_r)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} &= \frac{\widehat{\boldsymbol{\beta}}(L)\mathbf{G}(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}}, \\
\boldsymbol{\gamma}(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} &= \frac{\widehat{\boldsymbol{\gamma}}(L)\mathbf{G}(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}}, \\
\boldsymbol{\theta}(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} &= \frac{\widehat{\boldsymbol{\theta}}(L)\mathbf{G}(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}},
\end{aligned}$$

where $|\delta_k| < 1$ and the expansions of $\widehat{\boldsymbol{\beta}}(L), \widehat{\boldsymbol{\gamma}}(L), \widehat{\boldsymbol{\theta}}(L)$ contain only positive powers of L.[29] Note that $\{\delta_k\}$ collect all the inside poles of the primitives $\boldsymbol{\beta}(L)$, $\boldsymbol{\gamma}(L)$, and $\boldsymbol{\theta}(L)$, and also the inside poles of $\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}$ which are $\{\lambda_k\}$ and zeros.

The best response can be written as

$$\mathbb{E}_{it}[(\boldsymbol{\beta}(L) - \mathbf{I}_r)a_{it}] + \mathbb{E}_{it}[\boldsymbol{\xi}_{it}] + \mathbb{E}_{it}[\boldsymbol{\gamma}(L)a_t] = \mathbf{0}.$$

[29]Note that it is not necessary the case that those three fractions share the same inside poles. One can always set $\widehat{\boldsymbol{\beta}}(L)$, $\widehat{\boldsymbol{\gamma}}(L)$, and $\widehat{\boldsymbol{\theta}}(L)$ to remove the poles.

Denote $\boldsymbol{a}_{it} = \boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{x}_{it}$ as the equilibrium policy rule. It follows that $\boldsymbol{a}_t = \boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\boldsymbol{s}_{it}$. By the Wiener-Hopf prediction formula and Lemma A.1, we have

$$\mathbb{E}_{it}[(\boldsymbol{\beta}(L)-\mathbf{I}_r)\boldsymbol{a}_{it}] = (\boldsymbol{\beta}(L)-\mathbf{I}_r)\boldsymbol{h}(L)\boldsymbol{x}_{it} - \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left[\frac{\widehat{\boldsymbol{\beta}}(L)\boldsymbol{h}(L)\mathbf{M}(L)\mathbf{G}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it},$$

$$\mathbb{E}_{it}[\boldsymbol{\gamma}(L)\boldsymbol{a}_t] = \boldsymbol{\gamma}(L)\boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\boldsymbol{x}_{it} - \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left[\frac{\widehat{\boldsymbol{\gamma}}(L)\boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{G}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it},$$

$$\mathbb{E}_{it}[\boldsymbol{\xi}_{it}] = \boldsymbol{\theta}(L)\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\boldsymbol{x}_{it} - \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left[\frac{\widehat{\boldsymbol{\theta}}(L)\mathbf{G}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\boldsymbol{x}_{it}.$$

This has to be true for all the possible realizations of $\{\boldsymbol{x}_{it}\}$, which leads to the following fixed-point problem[30]

$$\begin{aligned}
&((\boldsymbol{\beta}(L)-\mathbf{I}_r)\otimes\mathbf{I}_n)\mathrm{vec}(\boldsymbol{h}'(L)) + (\boldsymbol{\gamma}(L)\otimes\boldsymbol{\rho}'_{xx}(L)^{-1}\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L))\mathrm{vec}(\boldsymbol{h}'(L)) = \\
&-(\mathbf{I}_r\otimes\boldsymbol{\rho}'_{xx}(L)^{-1}\mathbf{M}(L^{-1}))\mathrm{vec}(\boldsymbol{\theta}'(L)) + (\mathbf{I}_r\otimes\mathbf{B}'(L)^{-1}\mathbf{V}^{-1})\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\widehat{\boldsymbol{\theta}}'(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) \\
&+(\mathbf{I}_r\otimes\mathbf{B}'(L)^{-1}\mathbf{V}^{-1})\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\beta}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) \\
&+(\mathbf{I}_r\otimes\mathbf{B}'(L)^{-1}\mathbf{V}^{-1})\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\boldsymbol{\Lambda}\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\gamma}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right).
\end{aligned}$$

Multiplying $\mathbf{I}_r\otimes\boldsymbol{\rho}'_{xx}(L)$ to both sides yields

$$\begin{aligned}
&\mathbf{T}(L)\mathrm{vec}(\boldsymbol{h}'(L)) = -(\mathbf{I}_r\otimes\mathbf{M}(L^{-1}))\mathrm{vec}(\boldsymbol{\theta}'(L)) + \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{(\mathbf{I}_r\otimes\mathbf{B}(L^{-1}))}{\tau!(L-\delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\widehat{\boldsymbol{\theta}}'(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) \\
&+(\mathbf{I}_r\otimes\mathbf{B}(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left(\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\beta}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)+\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\boldsymbol{\Lambda}\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\gamma}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)\right),
\end{aligned} \tag{A.3}$$

where $\mathbf{T}(L)$ is

$$\Big((\boldsymbol{\beta}(L)-\mathbf{I}_r)\otimes\mathbf{M}(L^{-1})\mathbf{M}'(L) + (\boldsymbol{\gamma}(L)\otimes\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L))\Big).$$

Part II: determine the number of free constants. There are endogenous constants $\{\boldsymbol{h}^{(\tau)}(\delta_k)\}$ in condition (A.3) that remain to be determined. However, these constants may affect $\mathrm{vec}(\boldsymbol{h}'(L))$ in a linearly-dependent way. To count the degree of freedom of these constants properly, we rewrite the last term in condition (A.3) as follows:

$$\begin{aligned}
&(\mathbf{I}_r\otimes\mathbf{B}(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left(\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\beta}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)+\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\boldsymbol{\Lambda}\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\gamma}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)\right) \\
\equiv&(\mathbf{I}_r\otimes\mathbf{B}(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\mathbf{A}_\tau(\delta_k)\mathrm{vec}\begin{pmatrix}\boldsymbol{h}(\delta_k) & \boldsymbol{h}^{(1)}(\delta_k) & \dots & \boldsymbol{h}^{(\tau)}(\delta_k)\end{pmatrix}
\end{aligned}$$

[30]We use the following property of Kronecker product: $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}'\otimes\mathbf{A})\mathrm{vec}(\mathbf{B})$.

$$=(\mathbf{I}_r \otimes \mathbf{B}(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{\overline{\mathbf{A}}_\tau(\delta_k)}{\tau!(L-\delta_k)^{a_k-\tau}}\widetilde{\mathbf{A}}_\tau(\delta_k)\mathrm{vec}\left(\boldsymbol{h}(\delta_k) \quad \boldsymbol{h}^{(1)}(\delta_k) \quad \ldots \quad \boldsymbol{h}^{(\tau)}(\delta_k)\right), \tag{A.4}$$

where the constant matrices $\mathbf{A}_\tau(\delta_k)$ are simply constructed by collecting the corresponding exogenous terms, and $\mathbf{A}_\tau(\delta_k) = \overline{\mathbf{A}}_\tau(\delta_k)\widetilde{\mathbf{A}}_\tau(\delta_k)$ are obtained via the rank factorization. As a result, $\overline{\mathbf{A}}_\tau(\delta_k)$ is with full column rank and $\widetilde{\mathbf{A}}_\tau(\delta_k)$ is with full row rank. Accordingly, we can write (A.4) as the product of $\mathbf{D}_1(L)$ and $\boldsymbol{\psi}$, where

$$\mathbf{D}_1(L) \equiv (\mathbf{I}_r \otimes \mathbf{B}(L^{-1}))\left[\frac{\overline{\mathbf{A}}_0(\delta_1)}{0!(L-\delta_1)^{a_1}} \quad \ldots \quad \frac{\overline{\mathbf{A}}_{a_1-1}(\delta_1)}{(a_1-1)!(L-\delta_1)^1} \quad \ldots \quad \ldots \quad \frac{\overline{\mathbf{A}}_0(\delta_\ell)}{0!(L-\delta_\ell)^{a_\ell}} \quad \ldots \quad \frac{\overline{\mathbf{A}}_{a_\ell-1}(\delta_\ell)}{(a_\ell-1)!(L-\delta_\ell)^1}\right]$$

and

$$\boldsymbol{\psi} \equiv \left[\widetilde{\mathbf{A}}_0(\delta_1)\mathrm{vec}\left(\boldsymbol{h}(\delta_1)\right) \quad \ldots \quad \widetilde{\mathbf{A}}_{a_1-1}(\delta_1)\mathrm{vec}\left(\boldsymbol{h}(\delta_1) \quad \ldots \quad \boldsymbol{h}^{(a_1-1)}(\delta_1)\right) \quad \ldots\ldots \quad \widetilde{\mathbf{A}}_{a_\ell-1}(\delta_\ell)\mathrm{vec}\left(\boldsymbol{h}(\delta_\ell) \quad \ldots \quad \boldsymbol{h}^{(a_\ell-1)}(\delta_\ell)\right)\right]'.$$

We also define $\mathbf{D}_2(L)$ as

$$\mathbf{D}_2(L) \equiv -(\mathbf{I}_r \otimes \mathbf{M}(L^{-1}))\mathrm{vec}(\boldsymbol{\theta}'(L)) + \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{(\mathbf{I}_r \otimes \mathbf{B}(L^{-1}))}{\tau!(L-\delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\widehat{\boldsymbol{\theta}}'(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right).$$

Taking stock, the equilibrium policy rule $\boldsymbol{h}(L)$ needs to satisfy

$$\mathbf{T}(L)\mathrm{vec}(\boldsymbol{h}'(L)) = \mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L).$$

Here, by construction, $\mathbf{D}_1(L)$ is with full column rank, and its number of columns is given by N_ψ:

$$N_\psi = \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\mathrm{rank}(\mathbf{A}_\tau(\delta_k)).$$

A.6 Proof of Theorem 1

Proof. In order for $\boldsymbol{h}(L)$ to be a valid equilibrium policy rule, not only $\boldsymbol{h}(L)$ has to satisfy condition (3.13), but also it cannot contain any inside poles to guarantee it only use current or past information. We proceed in two steps: first, we specify the inside poles of $\boldsymbol{h}(L)$ that needs to be removed; second, we discuss whether these inside poles can be removed.

Part I: inside poles. To see the possible poles of $\boldsymbol{h}(L)$, we again consider the state-space representation of $\mathbf{M}(L)$ and $\mathbf{B}(L)$. Recall from the proof of Proposition 3.1, the signal process can be expressed with the state-space representation

$$\mathbf{M}(L) = \mathbf{H}\left(\mathbf{I}-\mathbf{F}L\right)^{-1}\boldsymbol{\Phi} + \boldsymbol{\Psi} = \frac{\mathbf{H}\mathrm{Adj}(\mathbf{I}-\mathbf{F}L)\boldsymbol{\Phi} + \boldsymbol{\Psi}\prod_{k=1}^{d}(1-\varkappa_k L)}{\prod_{k=1}^{d}(1-\varkappa_k L)} \equiv \frac{\mathbf{M}_1(L)}{\prod_{k=1}^{d}(1-\varkappa_k L)},$$
$$\mathbf{M}(L^{-1}) = \frac{\mathbf{H}\mathrm{Adj}(\mathbf{I}L-\mathbf{F})\boldsymbol{\Phi}L + \boldsymbol{\Psi}L^{v-d}\prod_{k=1}^{d}(L-\varkappa_k)}{L^{v-d}\prod_{k=1}^{d}(L-\varkappa_k)} \equiv \frac{\mathbf{M}_2(L)}{L^{v-d}\prod_{k=1}^{d}(L-\varkappa_k)},$$

where d is the number of non-zero eigenvalues of $\mathbf{F}$ and $\{\varkappa_k\}$ are the corresponding eigenvalues. By construction, all elements of $\mathbf{M}_1(L)$ and $\mathbf{M}_2(L)$ are polynomials in L.

Similarly,

$$\mathbf{B}(L) = \mathbf{I} + \mathbf{H}[\mathbf{I} - \mathbf{F}L]^{-1}\mathbf{FK}L = \frac{\mathbf{H}\mathrm{Adj}(\mathbf{I} - \mathbf{F}L)\mathbf{FK}L + \mathbf{I}\prod_{k=1}^{d}(1 - \varkappa_k L)}{\prod_{k=1}^{d}(1 - \varkappa_k L)} \equiv \frac{\mathbf{B}_1(L)}{\prod_{k=1}^{d}(1 - \varkappa_k L)},$$

$$\mathbf{B}(L^{-1}) = \frac{\mathbf{H}\mathrm{Adj}(\mathbf{I}L - \mathbf{F})\mathbf{FK} + \mathbf{I}L^{v-d}\prod_{k=1}^{d}(L - \varkappa_k)}{L^{v-d}\prod_{k=1}^{d}(L - \varkappa_k)} \equiv \frac{\mathbf{B}_2(L)}{L^{v-d}\prod_{k=1}^{d}(L - \varkappa_k)},$$

where all elements of $\mathbf{B}_1(L)$ and $\mathbf{B}_2(L)$ are polynomials in L.

We can now write $\mathbf{T}(L)$ as

$$\mathbf{T}(L) = \frac{(\boldsymbol{\beta}(L) - \mathbf{I}) \otimes (\mathbf{M}_2(L)\mathbf{M}_1'(L)) + \boldsymbol{\gamma}(L) \otimes \mathbf{M}_2(L)\boldsymbol{\Lambda}\mathbf{M}_1'(L)}{L^{v-d}\prod_{k=1}^{d}(1 - \varkappa_k L)(L - \varkappa_k)},$$

and the equilibrium policy rule in condition (3.13) can be expressed as

$$\begin{aligned}
\mathrm{vec}(\boldsymbol{h}'(L)) = &\left(\prod_{k=1}^{d}(1 - \varkappa_k L)\right)\Big((\boldsymbol{\beta}(L) - \mathbf{I}) \otimes (\mathbf{M}_2(L)\mathbf{M}_1'(L)) + \boldsymbol{\gamma}(L) \otimes \mathbf{M}_2(L)\boldsymbol{\Lambda}\mathbf{M}_1'(L)\Big)^{-1} \\
&\Bigg\{ - (\mathbf{I}_r \otimes \mathbf{M}_2(L^{-1}))\mathrm{vec}(\boldsymbol{\theta}'(L)) + \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{(\mathbf{I}_r \otimes \mathbf{B}_2(L^{-1}))}{\tau!(L - \delta_k)^{a_k-\tau}}\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\widehat{\boldsymbol{\theta}}'(L)}{\prod_{h\neq k}(L - \delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) \\
&+ (\mathbf{I}_r \otimes \mathbf{B}_2(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L - \delta_k)^{a_k-\tau}}\left(\mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\beta}}(L)}{\prod_{h\neq k}(L - \delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) + \mathrm{vec}\left(\left[\frac{\mathbf{G}'(L)\boldsymbol{\Lambda}\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\gamma}}(L)}{\prod_{h\neq k}(L - \delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)\right)\Bigg\} \quad \text{(A.5)}
\end{aligned}$$

Note that poles at $\{\delta_k\}$ are already removed by construction with the annihilation operator. As a result, the inside poles of $\boldsymbol{h}(L)$ only come from the inside roots of $\det[\mathbf{T}(L)]$, which are $\{\zeta_1, \zeta_2, \ldots, \zeta_{N_\zeta}\}$.

Part II: removing the inside poles. Let $\boldsymbol{\phi}(L)$ be a vector of rational functions in L with length $r \times n$, where its i-th element $\phi_i(L)$ is given by

$$\phi_i(L) \equiv \det\Big[\mathbf{T}_1(L) \;\; \ldots \;\; \mathbf{T}_{i-1}(L) \;\; \mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L) \;\; \mathbf{T}_{i+1}(L) \ldots \;\; \ldots \;\; \mathbf{T}_{rn}(L)\Big].$$

By Cramer's rule, the policy rule $\mathrm{vec}(\boldsymbol{h}'(L))$ that solves equation (3.13) is given by

$$\mathrm{vec}(\boldsymbol{h}'(L)) = \frac{\boldsymbol{\phi}(L)}{\det[\mathbf{T}(L)]}.$$

In order to remove the inside pole at ζ_j for $j \in \{1, \ldots, N_\zeta\}$, we need to solve for the vector of constants $\boldsymbol{\psi}$ such that ζ_j is a root of $\phi_i(L)$ for $i \in \{1, \ldots, rn\}$, and the existence and uniqueness of the equilibrium depends on existence and uniqueness of the vector of $\boldsymbol{\psi}$.

Denote $\mathcal{P}_j$ as the index set such that

$$\mathcal{P}_j = \{i \in \{1, \ldots, rn\} : \text{there exists some } \boldsymbol{\psi} \text{ such that } \phi_i(\zeta_j) \neq 0.\}$$

Throughout, we consider the relevant case where $\mathcal{P}_j \neq \emptyset$ for all $j \in \{1, \ldots, N_\zeta\}$, that is, the inside poles $\{\zeta_j\}$ are not automatically removed.

We first establish the following results: for $i \in \mathcal{P}_j$, when $\phi_i(\zeta_j) = 0$, then $\phi_k(\zeta_j) = 0$ for $k \neq i$. Note that $\phi_i(\zeta_j) = 0$ implies the following[31]

$$\mathbf{D}_1(\zeta_j)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_j) = \sum_{\tau \neq i} \kappa_\tau \mathbf{T}_\tau(\zeta_j),$$

where $\kappa_\tau \neq 0$ for some τ. As a result, for any $k \neq i$,

$$\begin{aligned}
\phi_k(\zeta_j) &= \det\Big[\mathbf{T}_1(\zeta_j) \;\; \ldots \;\; \underbrace{\mathbf{D}_1(\zeta_j)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_j)}_{k\text{-th column}} \;\; \ldots \;\; \mathbf{T}_{rn}(\zeta_j)\Big] \\
&= \sum_{\tau \neq i} \det\Big[\mathbf{T}_1(\zeta_j) \;\; \ldots \;\; \underbrace{\kappa_\tau \mathbf{T}_\tau(\zeta_j)}_{k\text{-th column}} \;\; \ldots \;\; \mathbf{T}_{rn}(\zeta_j)\Big] \\
&= \det\Big[\mathbf{T}_1(\zeta_j) \;\; \ldots \;\; \underbrace{\kappa_k \mathbf{T}_k(\zeta_j)}_{k\text{-th column}} \;\; \ldots \;\; \mathbf{T}_{rn}(\zeta_j)\Big] \\
&= 0,
\end{aligned}$$

where the last equality comes from the fact that $\det\big[\mathbf{T}(\zeta_j)\big] = 0$ as ζ_j is a root of $\det[\mathbf{T}(L)]$.

Next, we proceed to solve for the vector $\boldsymbol{\psi}$. For a collection of indexes $\boldsymbol{\iota} = \{\iota_1, \ldots, \iota_{N_\zeta}\}$ where $\iota_j \in \mathcal{P}_j$, define the pair of exogenous constant matrices $\{\mathbf{U}_1(\boldsymbol{\iota}), \mathbf{U}_2(\boldsymbol{\iota})\}$ as

$$\mathbf{U}_1(\boldsymbol{\iota})\boldsymbol{\psi} + \mathbf{U}_2(\boldsymbol{\iota}) = \Big[\phi_{\iota_1}(\zeta_1) \quad \ldots \quad \phi_{\iota_{N_\zeta}}(\zeta_{N_\zeta})\Big]'.$$

Let N_{rank} be given by

$$N_{\text{rank}} \equiv \max_{\boldsymbol{\iota}} \text{rank}(\mathbf{U}_1(\boldsymbol{\iota})).$$

We first consider the generic case where $N_{\text{rank}} = N_\zeta$.

- If $N_\zeta = N_\psi$, then there exists a unique equilibrium. Let $\boldsymbol{\iota}^* \in \operatorname{argmax}_{\boldsymbol{\iota}} \text{rank}(\mathbf{U}_1(\boldsymbol{\iota}))$. There exists a unique $\boldsymbol{\psi}^*$ such that

  $$\mathbf{U}_1(\boldsymbol{\iota}^*)\boldsymbol{\psi} + \mathbf{U}_2(\boldsymbol{\iota}^*) = \mathbf{0}. \tag{A.6}$$

  As a result, with $\mathbf{D}_1(L)\boldsymbol{\psi}^* + \mathbf{D}_2(L)$, we have

  $$\Big[\phi_{\iota_1}(\zeta_1) \quad \ldots \quad \phi_{\iota_{N_\zeta}}(\zeta_{N_\zeta})\Big]' = 0$$

  for any $\boldsymbol{\iota} = \{\iota_1, \ldots, \iota_{N_\zeta}\}$ where $\iota_j \in \{1, \ldots, rn\}$, as we have shown that $\phi_i(\zeta_j) = 0$ leads to $\phi_k(\zeta_j) = 0$ for $k \neq i$. By the same logic, it also follows that $\boldsymbol{\psi}^*$ does not depend on the the choice of ι^* given that is chosen among $\operatorname{argmax}_{\boldsymbol{\iota}} \text{rank}(\mathbf{U}_1(\boldsymbol{\iota}))$

  The equilibrium policy rule that satisfies conditon (3.13) can be expressed as

  $$\text{vec}(\boldsymbol{h}'(L)) = \mathbf{T}(L)^{-1}\Big(-\mathbf{D}_1(L)\mathbf{U}_1(\boldsymbol{\iota}^*)^{-1}\mathbf{U}_2(\boldsymbol{\iota}^*) + \mathbf{D}_2(L)\Big).$$

- If $N_\zeta < N_\psi$, then there exists an infinite number of solutions.

[31]Note that if $\{\mathbf{T}_1(\zeta_j), \ldots, \mathbf{T}_{i-1}(\zeta_j), \mathbf{T}_{i+1}(\zeta_j), \ldots, \mathbf{T}_{nr}(\zeta_j)\}$ are linearly dependent, then for any $\boldsymbol{\psi}$, $\phi_i(\zeta_j) = 0$, which implies the inside pole ζ_j is automatically removed.

To complete the picture, consider the case where $N_{\text{rank}} < N_\zeta$.

- If there exists any $\boldsymbol{\iota} = \{\iota_1, \ldots, \iota_{N_\zeta}\}$ where $\iota_j \in \mathcal{P}_j$ such that $\text{rank}(\mathbf{U}_1(\boldsymbol{\iota}), \mathbf{U}_2(\boldsymbol{\iota})) > \text{rank}(\mathbf{U}_1(\boldsymbol{\iota}))$, then there does not exist an equilibrium.
- If for any $\boldsymbol{\iota} = \{\iota_1, \ldots, \iota_{N_\zeta}\}$ where $\iota_j \in \mathcal{P}_j$, $\text{rank}(\mathbf{U}_1(\boldsymbol{\iota}), \mathbf{U}_2(\boldsymbol{\iota})) = \text{rank}(\mathbf{U}_1(\boldsymbol{\iota}))$, then there exists at least one equilibrium.

$\square$

Remark. To see why it is necessary to consider the case where $N_{\text{rank}} < N_\zeta$, we explore the following problem. Assume the best response is

$$\begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} s_t + \begin{bmatrix} \frac{1}{\rho_1+\rho_2} & 0 \\ 0 & \frac{1}{\theta_1+\theta_2} \end{bmatrix} \begin{bmatrix} \mathbb{E}_{it}[a_{t+1}^1] \\ \mathbb{E}_{it}[a_{t+1}^2] \end{bmatrix} + \begin{bmatrix} \frac{\rho_1\rho_2}{\rho_1+\rho_2} & 0 \\ 0 & \frac{\theta_1\theta_2}{\theta_1+\theta_2} \end{bmatrix} \begin{bmatrix} a_{t-1}^1 \\ a_{t-1}^2 \end{bmatrix},$$

and the signal is $x_t = s_t$. This is a perfect information economy, with $r = 2$, $m = 1$, and $n = 1$. What is special about this economy is that the two actions are independent of each other, similar to the examples discussed in Sims (2007), Onatski (2006), and Tan and Walker (2015).

Let $a_t^1 = h_1(L)s_t$ and $a_t^2 = h_2(L)s_t$ as the equilibrium policy rules, which need to satisfy the following condition

$$\underbrace{\begin{bmatrix} \frac{L^{-1}-(\rho_1+\rho_2)+\rho_1\rho_2 L}{\rho_1+\rho_2} & 0 \\ 0 & \frac{L^{-1}-(\theta_1+\theta_2)+\theta_1\theta_2 L}{\theta_1+\theta_2} \end{bmatrix}}_{\mathbf{T}(L)} \begin{bmatrix} h_1(L) \\ h_2(L) \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{L^{-1}}{(\rho_1+\rho_2)} & 0 \\ 0 & \frac{L^{-1}}{(\theta_1+\theta_2)} \end{bmatrix}}_{\mathbf{D}_1(L)} \underbrace{\begin{bmatrix} h_1(0) \\ h_2(0) \end{bmatrix}}_{\psi} + \underbrace{\begin{bmatrix} -1 \\ -1 \end{bmatrix}}_{\mathbf{D}_2(L)}.$$

The roots of $\mathbf{T}(L)$ are $\{\rho_1^{-1}, \rho_2^{-1}, \theta_1^{-1}, \theta_2^{-1}\}$. Consider the case where $|\rho_1| < 1$, $|\rho_2| < 1$, $|\theta_1| > 1$, and $|\theta_2| > 1$, that is, the inside roots are $\{\theta_1^{-1}, \theta_2^{-1}\}$. Clearly, since the the two actions are independent and both of the two inside roots are from a_t^2, there should be no solution. If we simply count the number of inside roots, $N_\zeta = 2$, and the number of constants, $N_\psi = 2$, it may seem to be the case that there is a unique solution. The flaw of this logic is that we still need to check N_{rank}.

With this parameterization, it is straightforward to verify that the index set $\mathcal{P}_1 = \mathcal{P}_2 = \{2\}$, and it follows $\boldsymbol{\iota} = \{2, 2\}$. Therefore, we have

$$\begin{bmatrix} \phi_2(\theta_1^{-1}) \\ \phi_2(\theta_2^{-1}) \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \frac{\theta_1^2-(\rho_1+\rho_2)\theta_1+\rho_1\rho_2}{(\rho_1+\rho_2)(\theta_1+\theta_2)} \\ 0 & \frac{\theta_2^2-(\rho_1+\rho_2)\theta_2+\rho_1\rho_2}{(\rho_1+\rho_2)(\theta_1+\theta_2)} \end{bmatrix}}_{\mathbf{U}_1(\boldsymbol{\iota})} \begin{bmatrix} h_1(0) \\ h_2(0) \end{bmatrix} + \underbrace{\begin{bmatrix} -\frac{\theta_1-(\rho_1+\rho_2)+\rho_1\rho_2\theta_1^{-1}}{\rho_1+\rho_2} \\ -\frac{\theta_2-(\rho_1+\rho_2)+\rho_1\rho_2\theta_2^{-1}}{\rho_1+\rho_2} \end{bmatrix}}_{\mathbf{U}_2(\boldsymbol{\iota})}$$

Note that $\text{rank}(\mathbf{U}_1(\boldsymbol{\iota})) < \text{rank}(\mathbf{U}_1(\boldsymbol{\iota}), \mathbf{U}_2(\boldsymbol{\iota}))$. Therefore, we can correctly conclude that there is no equilibrium.

A.7 Extension to Network Games

Consider the following network game. There are r different groups of agents, and there are a continuum of agents within each group. For agent i in group j, her best response is

$$a_{i,j,t} = \mathbb{E}_{i,j,t}[\theta_{i,j,t}] + \beta_j(L)a_{i,r,t} + \sum_{k=1}^{r} \gamma_{jk}(L)a_{kt},$$

where a_{kt} is the aggregate outcome of group k, $\beta_j(L)$ captures the PE consideration for group j, and $\gamma_{jk}(L)$ captures the GE dependence on different groups in the economy.

In addition to the heterogeneity in payoff structures, we also allow agents to have different information structures. Denote the signals observed by agent i in group j as

$$\boldsymbol{x}_{i,j,t} = \mathbf{M}_j(L)\boldsymbol{s}_{i,j,t}$$

Note that in this economy, the shocks at the group level cannot be washed out in aggregate. Denote $a_{i,j,t} = \boldsymbol{h}_j(L)\boldsymbol{x}_{i,j,t}$ as the equlibrium policy rule. It follows that the aggregate outcome in group j is

$$a_{jt} = \int a_{i,j,t} = \boldsymbol{h}_j(L)\mathbf{M}_j(L)\mathbf{\Lambda}\boldsymbol{s}_{i,j,t}.$$

Now consider the inference problem. We assume that all agents' signals follow ARMA processes and $\beta_j(L)$ and $\gamma_{jk}(L)$ are rational functions in L. We denote

$$\begin{aligned}
(\beta_j(L)-1)\mathbf{M}'_j(L^{-1})\mathbf{B}'_j(L^{-1})^{-1} &= \frac{\widehat{\beta}_j(L)\mathbf{G}_j(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}},\\
\gamma_{jk}(L)\mathbf{M}'_j(L^{-1})\mathbf{B}'_j(L^{-1})^{-1} &= \frac{\widehat{\gamma}_j(L)\mathbf{G}(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}},\\
\theta_j(L)\mathbf{M}'_j(L^{-1})\mathbf{B}'_j(L^{-1})^{-1} &= \frac{\widehat{\theta}_j(L)\mathbf{G}_j(L)}{\prod_{k=1}^{\ell}(L-\delta_k)^{a_k}},
\end{aligned}$$

where $\mathbf{G}_j(L)$ is constructed in a similar way as in Parallel to the proof of Proposition 3.1, and $\{\delta_k\}$ collect the eigenvalues of $\mathbf{F}_j - \mathbf{F}_j\mathbf{K}_j\mathbf{H}_j$ for all j and the inside poles from the payoff structures.

From the best response, the following has to hold for group j

$$(\beta_j(L)-1)\boldsymbol{\rho}'_{j,xx}(L)\boldsymbol{h}'_j(L) + \sum_{q=1}^{r}\gamma_{jq}(L)\mathbf{M}_j(L^{-1})\mathbf{\Lambda}\mathbf{M}'_q(L)\boldsymbol{h}'_q(L) = R_j(L),$$

where

$$\begin{aligned}
R_j(L) = &-\mathbf{M}_j(L^{-1}))\boldsymbol{\theta}'_j(L) + \\
&\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{\mathbf{B}_j(L^{-1})}{\tau!(L-\delta_k)^{a_k-\tau}}\left(\left[\frac{\mathbf{G}'_j(L)\widehat{\boldsymbol{\theta}}'_j(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k} + \left[\frac{\mathbf{G}'_j(L)\mathbf{M}'_j(L)\boldsymbol{h}'_j(L)\widehat{\beta}_j(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k} + \sum_{q=1}^{r}\left[\frac{\mathbf{G}'_j(L)\mathbf{\Lambda}\mathbf{M}'_q(L)\boldsymbol{h}'_q(L)\widehat{\gamma}_{jq}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right).
\end{aligned}$$

To set up the fixed-point problem, we define the following objects

$$\overline{\mathbf{M}}(L) \equiv \begin{bmatrix}\mathbf{M}_1(L) & & \\ & \ddots & \\ & & \mathbf{M}_r(L)\end{bmatrix}, \quad \overline{\beta}(L) \equiv \begin{bmatrix}\beta_1(L) & & \\ & \ddots & \\ & & \beta_r(L)\end{bmatrix}, \quad \overline{\gamma}(L) \equiv \begin{bmatrix}\gamma_{11}(L) & \dots & \gamma_{1r}(L)\\ \vdots & \ddots & \vdots \\ \gamma_{r1}(L) & \dots & \gamma_{rr}(L)\end{bmatrix}, \quad \overline{\mathbf{R}}(L) \equiv \begin{bmatrix}R_1(L)\\ \vdots \\ R_r(L)\end{bmatrix}.$$

By collecting the best responses for all r groups, we reach the functional equation of the equilibrium policy rule has to hold

$$\mathbf{T}(L)\begin{bmatrix} \boldsymbol{h}_1'(L) \\ \vdots \\ \boldsymbol{h}_r'(L) \end{bmatrix} = \mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L).$$

Here, the matrices $\mathbf{T}(L)$ is given by

$$\mathbf{T}(L) = \overline{\mathbf{M}}(L^{-1})(\overline{\boldsymbol{\beta}}(L) - \mathbf{I})\overline{\mathbf{M}}'(L) + \overline{\mathbf{M}}(L^{-1})(\overline{\boldsymbol{\gamma}}(L) \otimes \boldsymbol{\Lambda})\overline{\mathbf{M}}'(L),$$

and the matrices $\mathbf{D}_1(L)$ and $\mathbf{D}_2(L)$, and the endogenous constants $\boldsymbol{\psi}$ are defined as

$$\mathbf{D}_1(L)\boldsymbol{\psi} + \mathbf{D}_2(L) = \mathbf{R}(L)$$

with $\mathbf{D}_1(L)$ being with full column rank.

To determine the uniqueness and existence of the equilibrium policy rule, the procedure is identical to that in the proof of Theorem 1. Therefore, our results on the equilibrium policy rule and the finite-state representation extend to network games with incomplete information.

A.8 Proof of Corollary 3.1

Proof. The equilibrium outcome is given by

$$\boldsymbol{a}_{it} = \boldsymbol{h}(L)\mathbf{M}(L)\boldsymbol{s}_{it}.$$

Using the equilibrium policy representation (A.5) and the state-space representation of the signal process, we have

$$\begin{aligned}
\operatorname{vec}(\boldsymbol{h}(L)\mathbf{M}(L))' &= \frac{1}{\prod_{k=1}^{d}(1-\varkappa_k L)}(\mathbf{I}\otimes\mathbf{M}_1'(L))\operatorname{vec}(\boldsymbol{h}'(L)) \\
&= \Big((\boldsymbol{\beta}(L)-\mathbf{I})\otimes(\mathbf{M}_2(L)\mathbf{M}_1'(L)) + \boldsymbol{\gamma}(L)\otimes\mathbf{M}_2(L)\boldsymbol{\Lambda}\mathbf{M}_1'(L)\Big)^{-1} \\
&\quad \Bigg\{-(\mathbf{I}_r\otimes\mathbf{M}_2(L^{-1}))\operatorname{vec}(\boldsymbol{\theta}'(L)) + \sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{(\mathbf{I}_r\otimes\mathbf{B}_2(L^{-1}))}{\tau!(L-\delta_k)^{a_k-\tau}}\operatorname{vec}\left(\left[\frac{\mathbf{G}'(L)\widehat{\boldsymbol{\theta}}'(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) \\
&\quad + (\mathbf{I}_r\otimes\mathbf{B}_2(L^{-1}))\sum_{k=1}^{\ell}\sum_{\tau=0}^{a_k-1}\frac{1}{\tau!(L-\delta_k)^{a_k-\tau}}\left(\operatorname{vec}\left(\left[\frac{\mathbf{G}'(L)\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\beta}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right) + \operatorname{vec}\left(\left[\frac{\mathbf{G}'(L)\boldsymbol{\Lambda}\mathbf{M}'(L)\boldsymbol{h}'(L)\widehat{\boldsymbol{\gamma}}(L)}{\prod_{h\neq k}(L-\delta_h)^{a_h}}\right]^{(\tau)}_{L=\delta_k}\right)\right)\Bigg\}
\end{aligned}$$

Notice that the only possible auto-regressive parameters are from $\boldsymbol{\theta}(L)$ and the roots of $\det[\mathbf{T}(L)]$. □

A.9 Proof of Corollary 3.2

Proof. The proof follows from the main text. □

A.10 Proof of Corollary 3.3

Under Assumption 2, we can iterate individual's best response forward by the law of iterated expectations

$$\begin{aligned}
\boldsymbol{a}_{it} &= \mathbb{E}_{it}[\boldsymbol{\xi}_{it}] + \mathbb{E}_{it}[\beta(L)\boldsymbol{a}_{it}] + \mathbb{E}_{it}[\gamma(L)\boldsymbol{a}_t] \\
&= \mathbb{E}_{it}[(I-\beta(L))^{-1}(\boldsymbol{\xi}_{it} + \gamma(L)\boldsymbol{a}_t)] \\
&= \mathbb{E}_{it}[(I-\beta(L))^{-1}\boldsymbol{\xi}_{it}] + \mathbb{E}_{it}\left[(I-\beta(L))^{-1}\gamma(L)\mathbb{E}_{it}\left[(I-\beta(L))^{-1}\left(\int_j \boldsymbol{\xi}_{jt} + \gamma(L)\boldsymbol{a}_t\right)\right]\right] \\
&= \mathbb{E}_{it}[(I-\beta(L))^{-1}\boldsymbol{\xi}_{it}] + \mathbb{E}_{it}\left[(I-\beta(L))^{-1}\gamma(L)\mathbb{E}_{it}\left[(I-\beta(L))^{-1}\int_j \boldsymbol{\xi}_{jt}\right]\right] \\
&\quad + \mathbb{E}_{it}\left[(I-\beta(L))^{-1}\gamma(L)\mathbb{E}_{it}\left[(I-\beta(L))^{-1}\gamma(L)\left[(I-\beta(L))^{-1}\int_j \boldsymbol{\xi}_{jt}\right]\right]\right] + \ldots
\end{aligned}$$

Then, the variance of implied individual action is bounded by constant times the variance of fundamental as

$$\begin{aligned}
\mathbb{V}[\boldsymbol{a}_{it}] &\leq \mathbb{V}[(I-\beta(L))^{-1} + (I-\beta(L))^{-1}\gamma(L)(I-\beta(L))^{-1} + (I-\beta(L))^{-1}\gamma(L)(I-\beta(L))^{-1}\gamma(L)(I-\beta(L))^{-1} + \ldots)\boldsymbol{\xi}_{it}] \\
&= \mathbb{V}[(I-\beta(L))^{-1}\{I-\gamma(L)(I-\beta(L))^{-1}\}^{-1}\boldsymbol{\xi}_{it}] \\
&= \mathbb{V}[\{I-\beta(L)-\gamma(L)\}^{-1}\boldsymbol{\xi}_{it}] \\
&\leq \{I-\beta(1)-\gamma(1)\}^{-1}\mathbb{V}[\boldsymbol{\xi}_{it}](\{I-\beta(1)-\gamma(1)\}^{-1})'.
\end{aligned}$$

With exogenous signals, all the higher-order expectations are determined exogenously. Since the variance of a_{it} is bounded, it implies that there exists a unique solution. Aggregating across agents, the aggregate outcome can be expressed as

$$\boldsymbol{a}_t = \overline{\mathbb{E}}_t[(I-\beta(L))^{-1}\boldsymbol{\xi}_{it}] + \overline{\mathbb{E}}_t[(I-\beta(L))^{-1}\gamma(L)\boldsymbol{a}_t)] = \sum_{k=0}^{\infty}\overline{\mathbb{F}}_t^k[(\mathbf{I}-\boldsymbol{\beta}(L))^{-1}\boldsymbol{\xi}_{it}],$$

where $\overline{\mathbb{F}}^{k+1}[X] = \overline{\mathbb{E}}[(\mathbf{I}-\boldsymbol{\beta}(L))^{-1}\boldsymbol{\gamma}(L)\overline{\mathbb{F}}^k[X]]$. The proof is then completed by equating the higher-order expectations with the result from Corollary 3.1.

A.11 Proof of Proposition 4.1

The state equation and the observation equation are

$$\xi_t = \underbrace{\rho}_{\mathbf{F}}\ \xi_{t-1} + \underbrace{\sigma_\eta}_{\mathbf{\Phi}}\ \hat{\eta}_t, \qquad \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{H}}\ \xi_t + \underbrace{\begin{bmatrix} \sigma_\varepsilon & 0 \\ 0 & \sigma_u \end{bmatrix}}_{\mathbf{\Psi}}\begin{bmatrix} \hat{\varepsilon}_t \\ \hat{u}_{it} \end{bmatrix},$$

where $\eta_t = \sigma_\eta\hat{\eta}_t$, $\varepsilon_t = \hat{\varepsilon}_t$, and $u_{it} = \sigma_u\hat{u}_{it}$.

By Theorem 3.2, the prior variance of the state ξ_t and the Kalman gain matrix satisfies

$$\mathbf{P} = \mathbf{F}[\mathbf{P} - \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1}\mathbf{H}\mathbf{P}]\mathbf{F}' + \mathbf{\Phi}\mathbf{\Phi}', \qquad \mathbf{K} = \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1},$$

Denote $\kappa \equiv \mathbf{P}^{-1}$ as the prior precision about ξ_t, it is easy to verify that

$$\sigma_u^2\sigma_\varepsilon^2\kappa^2 = [(1-\rho^2)\sigma_u^2\sigma_\varepsilon^2 - \sigma_u^2\sigma_\eta^2 - \sigma_\varepsilon^2\sigma_\eta^2]\kappa + (\sigma_u^2+\sigma_\varepsilon^2)\sigma_\eta^2, \qquad \mathbf{K} = \begin{bmatrix} \sigma_u^2(\sigma_u^2\sigma_\varepsilon^2\kappa + \sigma_u^2+\sigma_\varepsilon^2)^{-1} & \sigma_\varepsilon^2(\sigma_u^2\sigma_\varepsilon^2\kappa + \sigma_u^2+\sigma_\varepsilon^2)^{-1} \end{bmatrix}.$$

Define $\tau_\varepsilon \equiv \frac{\sigma_\eta^2}{\sigma_\varepsilon^2}$, $\tau_u \equiv \frac{\sigma_\eta^2}{\sigma_u^2}$, and $\lambda \equiv \sigma_u^2\sigma_\varepsilon^2\kappa(\sigma_u^2\sigma_\varepsilon^2\kappa + \sigma_u^2+\sigma_\varepsilon^2)^{-1}\rho$, and it follows that

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

The fundamental representation is given by

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

Assuming $a_{it} = h_1(L)x_{it}^1 + h_2(L)x_{it}^2$, it follows that

$$a_t = (h_1(L)+h_2(L))\xi_t + h_1(L)\varepsilon_t.$$

By Proposition 3.1, we have

$$\begin{aligned}
\mathbb{E}_{it}[\xi_t] &= \begin{bmatrix} \frac{1}{1-\lambda L}\frac{\lambda\tau_\varepsilon}{(1-\rho\lambda)\rho} \\ \frac{1}{1-\lambda L}\frac{\lambda\tau_u}{(1-\rho\lambda)\rho} \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}, \\
\mathbb{E}_{it}[a_t] &= \begin{bmatrix} \frac{\lambda\tau_\varepsilon}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)}(h_1(L)+h_2(L)) - \frac{\lambda^2\tau_\varepsilon}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}(h_1(\lambda)+h_2(\lambda)) \\ \frac{\lambda\tau_u}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)}(h_1(L)+h_2(L)) - \frac{\lambda^2\tau_u}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}(h_1(\lambda)+h_2(\lambda)) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} \\
&+ \begin{bmatrix} \frac{\tau_u}{\tau_\varepsilon+\tau_u}h_1(L) + \frac{\tau_\varepsilon}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\tau_\varepsilon}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(\lambda-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \\ -\frac{\tau_u}{\tau_\varepsilon+\tau_u}h_1(L) + \frac{\tau_u}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\tau_u}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(\lambda-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}.
\end{aligned}$$

Notice that

$$\lambda + \frac{1}{\lambda} = \rho + \frac{1}{\rho} + \frac{\tau_\varepsilon+\tau_u}{\rho},$$

which leads to

$$\mathbb{E}_{it}[a_t] = \begin{bmatrix} \frac{\lambda\tau_\varepsilon}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)}(h_1(L)+h_2(L)) + \frac{\tau_u}{\tau_\varepsilon+\tau_u}h_1(L) + \frac{\tau_\varepsilon}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2\tau_\varepsilon}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \\ \frac{\lambda\tau_u}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)}(h_1(L)+h_2(L)) - \frac{\tau_u}{\tau_\varepsilon+\tau_u}h_1(L) + \frac{\tau_u}{\tau_\varepsilon+\tau_u}\frac{\lambda}{\rho}\frac{(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2\tau_u}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}.$$

The best response requires that

$$a_{it} = (1-\alpha)\mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[y_t],$$

which yields the following system of analytic functions

$$\mathbf{T}(L)\begin{bmatrix} h_1(L) \\ h_2(L) \end{bmatrix} = \underbrace{\begin{bmatrix} -\alpha\frac{\lambda^2\tau_\varepsilon}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)} \\ -\alpha\frac{\lambda^2\tau_\varepsilon}{\rho}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)} \end{bmatrix}}_{\mathbf{D}_1(L)} \underbrace{h_2(\lambda)}_{\psi} + \underbrace{\begin{bmatrix} \frac{1-\alpha}{1-\lambda L}\frac{\lambda\tau_\varepsilon}{(1-\rho\lambda)\rho} \\ \frac{1-\alpha}{1-\lambda L}\frac{\lambda\tau_u}{(1-\rho\lambda)\rho} \end{bmatrix}}_{\mathbf{D}_2(L)},$$

where $\mathbf{T}(L)$ is

$$\mathbf{T}(L) = \begin{bmatrix} 1-\alpha\frac{\lambda\tau_\varepsilon}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)} - \alpha\left(\frac{\tau_u}{\tau_\varepsilon+\tau_u} + \frac{\tau_\varepsilon}{\tau_\varepsilon+\tau_u}\frac{\frac{\lambda}{\rho}(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}\right) & -\alpha\frac{\lambda\tau_\varepsilon}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)} \\ -\alpha\frac{\lambda\tau_u}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)} - \alpha\left(-\frac{\tau_u}{\tau_\varepsilon+\tau_u} + \frac{\tau_u}{\tau_\varepsilon+\tau_u}\frac{\frac{\lambda}{\rho}(L-\rho)(1-\rho L)}{(1-\lambda L)(L-\lambda)}\right) & 1-\alpha\frac{\lambda\tau_u}{\rho}\frac{L}{(1-\lambda L)(L-\lambda)} \end{bmatrix}.$$

Note that

$$\det[\mathbf{T}(L)] = \frac{(1-\alpha)\lambda(L-\vartheta)(1-\vartheta L)}{\vartheta(1-\lambda L)(L-\lambda)},$$

where ϑ is the inside root of the determinant of $\mathbf{T}(L)$ given by

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

To remove the inside root of $\det[\mathbf{T}(L)]$, we choose $h_2(\lambda)$ such that the following object is zero when evaluated at $L=\vartheta$:

$$\det\left[\mathbf{D}_1(L)h_2(\lambda)+\mathbf{D}_2(L) \quad \mathbf{T}_2(L)\right] = \frac{1}{(1-\lambda L)(L-\lambda)}\left\{\frac{(1-\alpha)\lambda(L-\lambda)}{(1-\rho\lambda)\rho\tau_1} - \alpha\frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}(1-\rho L)h_2(\lambda)\right\},$$

which requires

$$h_2(\lambda) = \frac{(1-\alpha)(\vartheta-\lambda)}{\alpha\lambda(1-\rho\vartheta)}.$$

Using the Cramer's rule, we have

$$h_1(L) = \frac{\vartheta\tau_\varepsilon}{\rho(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \quad h_2(L) = \frac{(1-\alpha)\vartheta\tau_u}{\rho(1-\rho\vartheta)}\frac{1}{1-\vartheta L}.$$

The aggregate outcome is then

$$a_t = (h_1(L)+h_2(L))\xi_t + h_2(L)\varepsilon_t = \left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\xi_t + \frac{\tau_\varepsilon}{\tau_\varepsilon+(1-\alpha)\tau_u}\left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}\varepsilon_t,$$

where the second equality uses the fact that

$$\frac{\tau_\varepsilon+(1-\alpha)\tau_u}{\rho} = \frac{1}{\vartheta}+\vartheta-\frac{1}{\rho}-\rho.$$

Denote $\tau \equiv (1-\alpha)\tau_u+\tau_\varepsilon$. From the definition of ϑ, we have that ϑ satisfies

$$\vartheta+\frac{1}{\vartheta} = \rho+\frac{1+\tau}{\rho}. \tag{A.7}$$

It is straightforward to show that

$$\frac{\partial(\vartheta+\frac{1}{\partial\vartheta})}{\vartheta} = -\frac{1-\vartheta^2}{\vartheta^2} < 0,$$

and therefore the left-hand side of (A.7) is strictly decreasing in ϑ, and the right-hand side of (A.7) is strictly increasing in τ. Therefore, When $\tau \to 0$, the right-hand side approaches to $\rho+\frac{1}{\rho}$, which implies the upper bound for ϑ is ρ. When $\tau\to\infty$, the right-hand side approaches to infinity, which implies the lower bound for ϑ is 0.

A.12 Proof of Proposition 4.2

In the proof of Proposition 4.1, we have shown that ϑ is decreasing in $\tau \equiv (1-\alpha)\tau_u + \tau_\varepsilon$. Therefore, ϑ is increasing in α.

Since τ is increasing in τ_u and τ_ε, it follows that ϑ is decreasing in τ_u and τ_ε. By the implicit function theorem, condition (A.7) implies that

$$\frac{\partial \vartheta}{\partial \tau_u} = (1-\alpha)\frac{\partial \vartheta}{\partial \tau_\varepsilon}.$$

The volatility of the aggregate outcome driven by the common noise is given by

$$\mathbb{V}[a_t|\xi_t] = \tau_\varepsilon^{-1}\left(\frac{\tau_\varepsilon}{\tau_\varepsilon + (1-\alpha)\tau_u}\left(1 - \frac{\vartheta}{\rho}\right)\right)^2 \frac{1}{1-\vartheta^2}$$

As only the composite of $\widehat{\tau} \equiv (1-\alpha)\tau_u$ matters, it is sufficient to determine how $\mathbb{V}[a_t|\xi_t]$ depends on this composite. Again using condition (A.7), we have

$$\frac{\partial \vartheta}{\partial \widehat{\tau}} = \frac{\vartheta^2}{\rho(\vartheta^2 - 1)}.$$

The derivative of $\mathbb{V}[a_t|\xi_t]$ with respect to $\widehat{\tau}$ is

$$\begin{aligned}
\frac{\partial \mathbb{V}[a_t|\xi_t]}{\partial \widehat{\tau}} &= \frac{\tau_\varepsilon}{\rho^2}\left[-2\frac{1}{\tau^3}\frac{(\rho-\vartheta)^2}{1-\vartheta^2} + \frac{1}{\tau^2}\frac{-2(\rho-\vartheta)(1-\vartheta^2) + (\rho-\vartheta)^2 2\vartheta}{(1-\vartheta^2)^2}\frac{\partial \vartheta}{\partial \widehat{\tau}}\right] \\
&= -\frac{\tau_\varepsilon}{\rho^3}\frac{2(\rho-\vartheta)}{(1-\vartheta^2)^3\tau^3}\left((\rho-\vartheta)\rho(1-\vartheta^2)^2 - (1-\rho\vartheta)\vartheta^2\tau\right) \\
&= -\frac{\tau_\varepsilon}{\rho^3}\frac{2(\rho-\vartheta)}{(1-\vartheta^2)^3\tau^3}(\rho-\vartheta)^2(1-\rho\vartheta^3) < 0.
\end{aligned}$$

Therefore, $\mathbb{V}[a_t|\xi_t]$ is increasing in α.

Finally, the volatility of the aggregate outcome driven by the fundamental is given by

$$\mathbb{V}[a_t|\varepsilon_t] = \left(1 - \frac{\vartheta}{\rho}\right)^2 \frac{1+\rho\vartheta}{(1-\rho\vartheta)(1-\rho-\vartheta+\vartheta\rho)(1+\rho+\vartheta+\vartheta\rho)}$$

The derivative of $\mathbb{V}[a_t|\varepsilon_t]$ with respect to ϑ is

$$\frac{\partial \mathbb{V}[a_t|\varepsilon_t]}{\partial \vartheta} = -\frac{2\rho^2(1-\rho^2)^2(\rho-\vartheta)(1-\rho\vartheta^3)}{\rho^4(1-\rho\vartheta)^2(1-\rho^2)^2(1-\vartheta^2)^2} < 0.$$

Since ϑ is increasing in α, $\mathbb{V}[a_t|\varepsilon_t]$ is decreasing in α.

A.13 Proof of Lemma 4.1

Without informational frictions, guess that y_t is proportional to r_t. The consumption dynamics can be written as

$$c_{gt} = -(1-m_g)\sum_{k=0}^{\infty}(1-m_g)^k \mathbb{E}_t[r_{t+k}] + m_g\phi_g \sum_{k=0}^{\infty}(1-m_g)^k \mathbb{E}_t[y_{t+k}]$$

$$= -\frac{1-m_g}{1-(1-m_g)\rho} r_t + \frac{m_g \phi_g}{1-(1-m_g)\rho} y_t.$$

Since $y_t = \pi_1 c_{1t} + \pi_2 c_{2t}$, it follows that

$$\begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix} = \begin{bmatrix} -\frac{1-m_1}{1-(1-m_1)\rho} \\ -\frac{1-m_2}{1-(1-m_2)\rho} \end{bmatrix} r_t + \begin{bmatrix} \frac{m_1\phi_1\pi_1}{1-(1-m_1)\rho} & \frac{m_1\phi_1\pi_2}{1-(1-m_1)\rho} \\ \frac{m_2\phi_2\pi_1}{1-(1-m_2)\rho} & \frac{m_2\phi_2\pi_2}{1-(1-m_2)\rho} \end{bmatrix} \begin{bmatrix} c_{1t} \\ c_{2t} \end{bmatrix}.$$

Inverting the matrix leads to the desired result.

A.14 Proof of Proposition 4.3

The proof follows from Proposition 10 in Angeletos and Huo (2021).

A.15 Proof of Proposition 4.4

The output process can be written as

$$\begin{aligned} y_t &= \sum_g \pi_g \left(-(1-m_g) \sum_{k=0}^{\infty} (1-m_g)^k \overline{\mathbb{E}}_t[r_{t+k}] + m_g \phi_g \sum_{k=0}^{\infty} (1-m_g)^k \overline{\mathbb{E}}_t[y_{t+k}] \right) \\ &= -\sum_g \frac{\pi_g(1-m_g)}{1-(1-m_g)\rho} \overline{\mathbb{E}}_t[r_t] + \sum_g \pi_g m_g \phi_g \overline{\mathbb{E}}_t \left[\frac{1}{1-(1-m_g)L^{-1}} y_t \right], \end{aligned}$$

where we have used the assumption of common information structure. Denote policy function as $y_t = h(L) r_t$. The average forecasts are

$$\begin{aligned} \overline{\mathbb{E}}_t[r_t] &= \left(1 - \frac{\lambda}{\rho}\right) \frac{1}{1-\lambda L} \frac{1}{1-\rho L} \eta_t \\ \overline{\mathbb{E}}_{1t}\left[\frac{1}{1-(1-m_g)L^{-1}} y_t \right] &= \left(\frac{h(L)L^2}{(1-\rho L)(L-\lambda)(L-(1-m_g))} - \frac{h(\lambda)\lambda^2}{(1-\rho\lambda)(L-\lambda)(\lambda-(1-m_g))} \right. \\ &\quad \left. - \frac{h(1-m_g)(1-m_g)^2}{(1-\rho(1-m_g))(1-m_g-\lambda)(L-(1-m_g))} \right) \frac{\tau\lambda}{\rho} \frac{1-\rho L}{1-\lambda L} \frac{1}{1-\rho L} \eta_t. \end{aligned}$$

The equilibrium condition requires that

$$\begin{aligned} h(L) = &-\sum_g \frac{\pi_g(1-m_g)}{1-(1-m_g)\rho} \left(1-\frac{\lambda}{\rho}\right) \frac{1}{1-\lambda L} + \sum_g \pi_g m_g \phi_g \left(\frac{h(L)L^2}{(1-\rho L)(L-\lambda)(L-(1-m_g))} - \frac{h(\lambda)\lambda^2}{(1-\rho\lambda)(L-\lambda)(\lambda-(1-m_g))} \right. \\ &\left. - \frac{h(1-m_g)(1-m_g)^2}{(1-\rho(1-m_g))(1-m_g-\lambda)(L-(1-m_g))} \right) \frac{\tau\lambda}{\rho} \frac{1}{1-\lambda L}. \end{aligned}$$

The endogenous persistence ϑ is the inverse of the outside root of the following equation

$$C(L) = 1 - \frac{\tau\lambda}{\rho} \sum_g \pi_g m_g \phi_g \frac{L}{(L-\lambda)1-\lambda L)(1-(1-m_g)L^{-1})},$$

where the numerator is proportional to

$$T(L)=\sum_g \pi_g \frac{m_g\phi_g}{1-(1-m_g)L^{-1}}-\frac{(1-\rho L)(L-\rho)+\tau L}{\tau L}.$$

The inside roots of $T(L)$ can be removed by chosen $h(\lambda)$ and $h(1-m_g)$ properly. This leads to that

$$y_t=\left(1-\frac{\vartheta}{\rho}\right)\frac{1}{1-\vartheta L}y_t^*.$$

Rearranging $T(L)$, we have

$$T(L)=\frac{\tau L^2\Big(\pi_1 m_1\phi_1(L-(1-m_2))+\pi_2 m_2\phi_2(L-(1-m_1))\Big)-((1-\rho L)(L-\rho)+\tau L)(L-(1-m_1))(L-(1-m_2))}{(L-(1-m_1))(L-(1-m_2))L}$$

Let $D(L)$ denote the numerator of $T(L)$. Recall that $\pi_1\phi_1+\pi_2\phi_2=1$, it follows that

$$\frac{\partial D(L)}{\partial\phi_1}=-m_1\pi_1(\mathrm{m}_1-\mathrm{m}_2)\tau L^2.$$

Also note that

$$\begin{aligned}
D(\lambda)&=\frac{\tau}{\lambda^2}\Big(\pi_1 m_1\phi_1(\lambda^{-1}-(1-m_2))+\pi_2 m_2\phi_2(\lambda^{-1}-(1-m_1))\Big)>0\\
D(1)&=-\mathrm{m}_1\mathrm{m}_2(1-\rho)^2<0.
\end{aligned}$$

Therefore, $\vartheta\in(\lambda,1)$ and $D(z)$ is decreasing in the neighborhood of ϑ^{-1}. When $\mathrm{m}_1>\mathrm{m}_2$, $\frac{\partial D(L)}{\partial\phi_1}|_{L=\vartheta^{-1}}<0$. It follows that ϑ is increasing in ϕ_1.

Notice that when $m_1=m_2=m$, $T(L)$ becomes

$$T(L)=\frac{m}{1-(1-m)L^{-1}}-\frac{(1-\rho L)(L-\rho)+\tau L}{\tau L}.$$

Therefore, ϕ_g is irrelevant in determining ϑ.

A.16 Proof of Proposition 4.5

Due to that $\phi_2=0$, the consumption dynamics for the low MPC group is given by

$$c_{2,t}=-(1-m_2)\sum_{k=0}^{\infty}(1-m_2)^k\overline{\mathbb{E}}_{2,t}[r_{t+k}]=-\frac{1-m_2}{1-(1-m_2)\rho}\overline{\mathbb{E}}_{2,t}[r_t].$$

That is, the consumption dynamics is proportional to the first-order expectation of r_t:

$$\begin{aligned}
c_{2t}&=\left(1-\frac{\lambda_2}{\rho}\right)\frac{1}{1-\lambda_2 L}\left(-\frac{1-m_2}{1-(1-m_2)\rho}r_t\right)\\
&=\left(1-\frac{\lambda_2}{\rho}\right)\frac{1}{1-\lambda_2 L}c_{2t}^*
\end{aligned}$$

$$= c_{2t}^* - \frac{\lambda_2}{\rho}\frac{1}{1-\lambda_2 L}c_2^*\eta_t.$$

Turn to consumers in group 1. Their problem can be written as

$$\begin{aligned}
c_{1t} &= -\frac{1-m_1}{1-(1-m_1)\rho}\overline{\mathbb{E}}_{1t}[r_t] + m_1\phi_1\sum_{k=0}^{\infty}(1-m_1)^k\pi_2\overline{\mathbb{E}}_{1t}[c_{2,t+k}] + m_1\phi_1\sum_{k=0}^{\infty}(1-m_1)^k\pi_1\overline{\mathbb{E}}_{1t}[c_{1,t+k}] \\
&= \frac{1}{1-(1-m_1)\rho}\left(-(1-m_1)+m_1\phi_1\pi_2 c_2^*\right)\overline{\mathbb{E}}_{1t}[r_t] - m_1\phi_1\pi_2 c_2^*\frac{\lambda_2}{\rho}\frac{1}{1-(1-m_1)\lambda_2}\overline{\mathbb{E}}_{1t}\left[\frac{1}{1-\lambda_2 L}\eta_t\right] + \overline{\mathbb{E}}_{1t}\left[\frac{m_1\phi_1\pi_1}{1-(1-m_1)L^{-1}}c_{1t}\right].
\end{aligned} \tag{A.8}$$

Denote individual policy function as $c_{i,1,t} = h(L)x_{i,1,t}$. The average forecasts are

$$\begin{aligned}
\overline{\mathbb{E}}_{1t}[r_t] &= \left(1-\frac{\lambda_1}{\rho}\right)\frac{1}{1-\lambda_1 L}\frac{1}{1-\rho L}\eta_t \\
\overline{\mathbb{E}}_{1t}\left[\frac{1}{1-\lambda_2 L}\eta_t\right] &= \left(1-\frac{\lambda_1}{\rho}\right)\frac{1-\rho\lambda_1}{1-\lambda_2\lambda_1}\frac{1}{1-\lambda_1 L}\frac{1}{1-\lambda_2 L}\eta_t \\
\overline{\mathbb{E}}_{1t}\left[\frac{1}{1-(1-m_1)L^{-1}}c_{1t}\right] &= \left(\frac{h(L)L^2}{(1-\rho L)(L-\lambda_1)(L-(1-m_1))} - \frac{h(\lambda_1)\lambda_1^2}{(1-\rho\lambda_1)(L-\lambda_1)(\lambda_1-(1-m_1))}\right. \\
&\quad \left. - \frac{h(1-m_1)(1-m_1)^2}{(1-\rho(1-m_1))(1-m_1-\lambda)(L-(1-m_1))}\right)\frac{\tau_1\lambda_1}{\rho}\frac{1-\rho L}{1-\lambda_1 L}\frac{1}{1-\rho L}\eta_t.
\end{aligned}$$

Notice that in the equation (A.8), the first two terms are exogenous and require only first-order expectation, and only the last term involves higher-order expectations. By Theorem 1 and Lemma 3.2, the endogenous persistence is the inverse of the outside root of the following equation

$$\begin{aligned}
C(L) &= \begin{bmatrix} \frac{L}{1-\rho} & \tau_1^{-\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{1-\rho L} \\ \tau_1^{-\frac{1}{2}} \end{bmatrix} - \frac{L}{L-\rho}\frac{1}{1-\rho L}\frac{m_1\phi_1\pi_1}{1-(1-m_1)L^{-1}} \\
&= \frac{L+\tau_1^{-1}(1-\rho L)(L-\rho) - L\frac{m_1\phi_1\pi_1}{1-(1-m_1)L^{-1}}}{(1-\rho L)(L-\rho)},
\end{aligned}$$

where the numerator is proportional to

$$T(L) = \frac{m_1\phi_1\pi_1}{1-(1-m_1)L^{-1}} - \frac{(1-\rho L)(L-\rho)+\tau_1 L}{\tau_1 L}.$$

The two inside roots can be removed by chosen $h(\lambda_1)$ and $h(1-m_1)$ accordingly. This leads to that

$$c_{1t} - c_{1t}^* = -c_1^*\frac{\vartheta}{\rho}\frac{1}{1-\vartheta L}\eta_t - m_1\phi_1\pi_2\frac{1-\frac{\vartheta}{\lambda_2}}{T(\lambda_2^{-1})(1-(1-m_1)\lambda_2)}\frac{1}{1-\vartheta L}c_2^*\frac{\lambda_2}{\rho}\frac{1}{1-\lambda_2 L}\eta_t.$$

A.17 Proof of Proposition 4.6

Note that $x_{m(i,t),t}^1 = a_i + \varepsilon_{m(i,t)t}$, the signal process can be equivalently rewritten as

$$x_{it}^1 = \kappa_{m(i,t)} + \varepsilon_{it}, \quad \widehat{x}_{it}^2 = x_{it}^2 - \kappa_i = \xi_t + \varepsilon_{m(i,t),t} + u_{it}, \quad \xi_t = \rho\xi_{t-1} + \eta_t.$$

Denote the policy rule using this transformed signals as

$$a_{it} = g_0\kappa_i + g_1(L)x^1_{it} + g_2(L)\widehat{x}^2_{it}.$$

In the end, the policy rule using the original signals can be found by

$$h_a = g_0 - g_2(1), \qquad h_1(L) = g_1(L), \qquad h_2(L) = g_2(L).$$

Note that the two signals are independent of each other, and we can find the fundamental representation for each of them separately. The fundamental representation for $\widehat{x}^2_{it}$ is

$$B(L) = \frac{1-\lambda L}{1-\rho L}, \qquad V = \frac{\rho(\sigma^2_\varepsilon + \sigma^2_u)}{\lambda\sigma^2_\eta},$$

where

$$\lambda = \frac{1}{2}\left[\rho + \frac{1}{\rho} + \frac{\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)}\right)^2 - 4}\,\right]. \tag{A.9}$$

Denote $j \equiv m(i,t)$ as the index of agent i's current match. The prediction of a_{jt} is

$$\mathbb{E}_{it}[a_{jt}] = \mathbb{E}_{it}[g_0\kappa_j + g_1(L)(\kappa_{m(j,t)} + \varepsilon_{jt}) + g_2(L)(u_{jt} + \varepsilon_{m(j,t),t} + \xi_t)],$$

where

$$\mathbb{E}_{it}[\kappa_j] = \frac{\sigma^2_\kappa}{\sigma^2_\kappa+\sigma^2_\varepsilon}x^1_{it}, \quad \mathbb{E}_{it}[u_{jt}] = 0, \quad \mathbb{E}_{it}[\varepsilon_{jt}] = \frac{\lambda\sigma^2_\varepsilon\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)}\frac{1-\rho L}{1-\lambda L}\widehat{x}^2_{it}, \quad \mathbb{E}_{it}[g_2(L)\xi_t] = \left(\frac{Lg_2(L)}{L-\lambda} - \frac{\lambda(1-\rho L)g_2(\lambda))}{(1-\rho\lambda)(L-\lambda)}\right)\frac{V^{-1}}{1-\lambda L}\widehat{x}^2_{it},$$

$$\mathbb{E}_{it}[\kappa_{m(j,\tau),\tau}] = \kappa_i, \quad \mathbb{E}_{it}[\varepsilon_{m(j,\tau),\tau}] = \frac{\sigma^2_\varepsilon}{\sigma^2_\kappa+\sigma^2_\varepsilon}x^1_{it} \quad \text{if } \tau = t\text{, otherwise } 0.$$

The best response requires that

$$\begin{aligned}
&g_0\kappa_i + g_1(L)x^1_{it} + g_2(L)\widehat{x}^2_{it} \\
=&\kappa_i + \alpha\left[g_0\frac{\sigma^2_\kappa}{\sigma^2_\kappa+\sigma^2_\varepsilon}x^1_{it} + g_1(0)\kappa_i + g_1(0)\frac{\lambda\sigma^2_\varepsilon\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)}\frac{1-\rho L}{1-\lambda L}\widehat{x}^2_{it} + \left(\frac{Lg_2(L)}{L-\lambda} - \frac{\lambda(1-\rho L)g_2(\lambda))}{(1-\rho\lambda)(L-\lambda)}\right)\frac{V^{-1}}{1-\lambda L}\widehat{x}^2_{it} + g_2(0)\frac{\sigma^2_\varepsilon}{\sigma^2_\kappa+\sigma^2_\varepsilon}x^1_{it}\right],
\end{aligned}$$

which leads to

$$g_0 = 1 + \alpha g_1(0), \qquad g_1(0) = \alpha g_0\frac{\sigma^2_\kappa}{\sigma^2_\kappa+\sigma^2_\varepsilon} + \alpha g_2(0)\frac{\sigma^2_\varepsilon}{\sigma^2_\kappa+\sigma^2_\varepsilon},$$
$$g_2(L) = \alpha g_1(0)\frac{\lambda\sigma^2_\varepsilon\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)}\frac{1-\rho L}{1-\lambda L} + \alpha\left(\frac{Lh_2(L)}{L-\lambda} - \frac{\lambda(1-\rho L)g_2(\lambda))}{(1-\rho\lambda)(L-\lambda)}\right)\frac{V^{-1}}{1-\lambda L}.$$

The third equation can be written as

$$-(L-\vartheta)\left(L - \frac{1}{\vartheta}\right)g_2(L) = \alpha g_1(0)\frac{\sigma^2_\varepsilon\sigma^2_\eta}{\rho(\sigma^2_\varepsilon+\sigma^2_u)}(1-\rho L)(L-\lambda) - \alpha\frac{V^{-1}(1-\rho L)g_2(\lambda)}{(1-\rho\lambda)},$$

where

$$\vartheta = \frac{1}{2}\left[\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\eta^2}{\rho(\sigma_\varepsilon^2+\sigma_u^2)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\eta^2}{\rho(\sigma_\varepsilon^2+\sigma_u^2)}\right)^2 - 4}\right].$$

Use $g_2(\lambda)$ to removes the inside root ϑ, we have

$$g_1(L) = g_1(0) = \frac{\alpha}{1-\alpha^2+\frac{\sigma_\varepsilon^2}{\sigma_\kappa^2}\left(1-\alpha^2\frac{\vartheta}{\rho}\frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_\varepsilon^2+\sigma_u^2}\right)}, \quad g_0 = 1+\alpha g_1(0), \quad g_2(L) = \frac{\alpha\vartheta g_1(0)\sigma_\varepsilon^2\sigma_\eta^2}{\rho(\sigma_\varepsilon^2+\sigma_u^2)}\frac{1-\rho L}{1-\vartheta L}$$

By aggregation, the aggregate outcome a_t is

$$a_t = g_2(L)\xi_t = \frac{\varphi}{1-\vartheta L}\eta_t, \quad \text{where} \quad \varphi = \frac{\alpha^2\vartheta}{\rho}\frac{\sigma_\eta^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_u^2}\left(1-\alpha^2+\frac{\sigma_\varepsilon^2}{\sigma_\kappa^2}\left(1-\alpha^2\frac{\vartheta}{\rho}\frac{\sigma_\eta^2\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_u^2}\right)\right)^{-1}.$$

A.18 Proof of Proposition 4.7

First consider the case with rational expectations. By construction, rational expectations imply that $K_{\text{BGMS}} = 0$. The forecast about future aggregate outcome is given by

$$\overline{\mathbb{E}}_t[a_t] = \varphi\frac{\rho-\lambda}{\rho}\frac{1-\rho\lambda}{1-\vartheta\lambda}\frac{1}{(1-\lambda L)(1-\vartheta L)}\eta_t,$$

where λ is defined in equation (A.9).

Denote $\delta \equiv \frac{\rho-\lambda}{\rho}\frac{1-\rho\lambda}{1-\vartheta\lambda}$. Note that $\vartheta < \rho$ when $\alpha < 1$, which implies that $\delta \in (0,1)$ when $\alpha < 1$. The aggregate forecast error and the aggregate forecast revision are

$$a_{t+1} - \overline{\mathbb{E}}_t[a_{t+1}] = a_{t+1} - \vartheta\overline{\mathbb{E}}_t[a_t] = \varphi\eta_{t+1} + \vartheta\varphi\frac{1}{1-\vartheta L}\left(1-\frac{\delta}{1-\lambda L}\right)\eta_t,$$

$$\overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}] = \vartheta\overline{\mathbb{E}}_t[a_t] - \vartheta^2\overline{\mathbb{E}}_{t-1}[a_{t-1}] = \varphi\vartheta\delta\frac{1}{1-\lambda L}\eta_t.$$

To compute K_{CG}, we need to compute the variance of the forecast revision and the covariance between forecast error and the forecast revision, which are

$$\mathbb{V}\left(\overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}]\right) = \left(\varphi\vartheta\frac{\rho-\lambda}{\rho}\frac{1-\rho\lambda}{1-\vartheta\lambda}\right)^2\frac{1}{1-\lambda^2},$$

$$\mathbb{COV}\left(a_{t+1} - \overline{\mathbb{E}}_t[a_{t+1}], \overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}]\right) = \varphi^2\vartheta^2\mathbb{COV}\left(\frac{1}{1-\vartheta L}\left(1-\frac{\delta}{1-\lambda L}\right)\eta_t, \delta\frac{1}{1-\lambda L}\eta_t\right).$$

This leads to

$$K_{\text{CG}} = \frac{\mathbb{COV}\left(a_{t+1} - \overline{\mathbb{E}}_t[a_{t+1}], \overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}]\right)}{\mathbb{V}\left(\overline{\mathbb{E}}_t[a_{t+1}] - \overline{\mathbb{E}}_{t-1}[a_{t+1}]\right)} = \lambda\frac{-\rho(1-\lambda^2)\vartheta + 1 + \rho^2 - 2\rho\lambda}{(\rho-\lambda)(1-\rho\lambda)(1-\lambda\vartheta)}$$

To show that $K_{CG} > 0$, we use the property that $\vartheta < \rho$, which leads to

$$K_{CG} > \lambda \frac{-\rho(1-\lambda^2)\rho + 1 + \rho^2 - 2\rho\lambda}{(\rho-\lambda)(1-\rho\lambda)(1-\lambda\vartheta)} = \lambda \frac{(1-\rho\lambda)^2}{(\rho-\lambda)(1-\rho\lambda)(1-\lambda\vartheta)} > 0.$$

To verify that K_{CG} is monotonically decreasing in α, it is sufficient to show that

$$\frac{\partial K_{CG}}{\partial \vartheta} = -\frac{\lambda}{(1-\lambda\vartheta)^2} < 0,$$

and the desired result follows from the fact that ϑ is monotonically increasing in α.

Second, consider the case with heterogeneous prior. In this case, we have

$$f_1 a_i + f_2 a_{m(i,t)} + f_3 \xi_t = a_i + \alpha(f_1 a_{m(i,t)} + f_2(a_i + \xi_t) + f_3 \xi_t).$$

By the method of undetermined coefficients, we have

$$f_1 = 1 + \alpha f_2, \quad f_2 = \alpha f_1, \quad f_3 = \alpha(f_2 + f_3),$$

which leads to

$$f_1 = \frac{1}{1-\alpha^2}, \quad f_2 = \frac{\alpha}{1-\alpha^2}, \quad f_3 = \frac{\alpha^2}{(1-\alpha)(1-\alpha^2)}.$$

The individual's forecast about aggregate output is given by

$$\mathbb{E}_{it}[a_t] = (f_2 + f_3)\xi_t = \frac{f_3}{\alpha}\xi_t.$$

Note that this forecast is not the rational forecast. The forecast error and forecast revision are

$$a_{t+1} - \mathbb{E}_{it}[a_{t+1}] = f_3 \xi_{t+1} - \rho \frac{f_3}{\alpha}\xi_t, \quad \mathbb{E}_{it}[a_{t+1}] - \mathbb{E}_{it-1}[a_{t+1}] = \rho \frac{f_3}{\alpha}\xi_t - \rho^2 \frac{f_3}{\alpha}\xi_{t-1}.$$

It is straightforward to show that

$$\mathbb{V}(\mathbb{E}_{it}[a_{t+1}] - \mathbb{E}_{it-1}[a_{t+1}]) = \left(\rho \frac{f_3}{\alpha}\right)^2, \quad \mathbb{COV}\left(a_{t+1} - \mathbb{E}_{it}[a_{t+1}], \mathbb{E}_{it}[a_{t+1}] - \mathbb{E}_{i,t-1}[a_{t+1}]\right) = (\rho f_3)^2 \frac{\alpha - 1}{\alpha^2}.$$

In this case, the individual forecast and the average forecast are identical. Therefore,

$$K_{CG} = K_{BGMS} = \alpha - 1.$$

A.19 Proof of Proposition 4.8

We proceed by a guess-and-verify approach. Suppose that the aggregate outcome a_t in the equilibrium has a finite-state representation. If this is the case, then there exists a state-space representation for an individual agent. Let $\boldsymbol{\theta}_t$ be the vector of underlying states, and a_t is related with $\boldsymbol{\theta}_t$ through a matrix $\mathbf{A}$: $a_t = \mathbf{A}\boldsymbol{\theta}_t$. The signal is related to the underlying state via

$x_{it} = \mathbf{H}\boldsymbol{\theta}_t + u_{it}$. The optimal forecast for the underlying state is

$$\mathbb{E}_{it}[\boldsymbol{\theta}_t] = \mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t] + \mathbf{K}(x_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t]),$$

where $\mathbf{K}$ is the corresponding Kalman gain matrix. It follows that

$$\mathbb{E}_{it}[a_t] = \mathbb{E}_{i,t-1}[a_t] + \mathbf{AK}(x_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t]).$$

Now consider the diagnostic expectations

$$\widetilde{\mathbb{E}}_{it}[\boldsymbol{\theta}_t] = \mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t] + (1+\mu)\mathbf{K}(x_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t]),$$

which implies that

$$\mathbf{A}\widetilde{\mathbb{E}}_{it}[\boldsymbol{\theta}_t] = \mathbf{A}\mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t] + (1+\mu)\mathbf{AK}(x_{it} - \mathbf{H}\mathbb{E}_{i,t-1}[\boldsymbol{\theta}_t]),$$

and

$$\widetilde{\mathbb{E}}_{it}[a_t] = \mathbb{E}_{i,t-1}[a_t] + (1+\mu)\mathbf{AK}(x_{it} - \mathbf{H}\mathbb{E}_{it-1}[\boldsymbol{\theta}_t]) = (1+\mu)\mathbb{E}_{it}[a_t] - \mu\mathbb{E}_{i,t-1}[a_t].$$

Let $a_{it} = h(L)x_{it}$ as the equilibrium policy rule. The forecasts about the aggregate outcome is then given by

$$\mathbb{E}_{it}[a_t] = \frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L)L - h(\lambda)\lambda\frac{1-\rho L}{1-\rho\lambda}\right)x_{it}, \quad \mathbb{E}_{it}[a_{t+1}] = \frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L) - h(\lambda)\frac{1-\rho L}{1-\rho\lambda}\right)x_{it},$$

which leads to

$$\widetilde{\mathbb{E}}_{it}[\xi_t] = \left(1-\frac{\lambda}{\rho}\right)\frac{1+\mu-\mu\rho L}{1-\lambda L}x_{it}, \quad \widetilde{\mathbb{E}}_{it}[a_t] = \frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L)((1+\mu)L-\mu L) - h(\lambda)((1+\mu)\lambda-\mu L)\frac{1-\rho L}{1-\rho\lambda}\right)x_{it}.$$

The best response can be written as

$$h(L)x_{it} = (1-\alpha)\left[\left(1-\frac{\lambda}{\rho}\right)\frac{1+\mu-\mu\rho L}{1-\lambda L}x_{it}\right] + \alpha\left[\frac{\lambda}{\rho\sigma^2(1-\lambda L)(L-\lambda)}\left(h(L)L - h(\lambda)((1+\mu)\lambda-\mu L)\frac{1-\rho L}{1-\rho\lambda}\right)x_{it}\right],$$

which leads to the following fixed-point problem

$$\left((1-\lambda L)(L-\lambda) - \alpha\frac{\lambda}{\rho\sigma^2}L\right)h(L) = (1-\alpha)\left(1-\frac{\lambda}{\rho}\right)(1+\mu-\mu\rho L)(L-\lambda) - \alpha\frac{\lambda}{\rho\sigma^2}h(\lambda)((1+\mu)\lambda-\mu L)\frac{1-\rho L}{1-\rho\lambda}$$

The roots of $\left((1-\lambda L)(L-\lambda) - \alpha\frac{\lambda}{\rho\sigma^2}L\right)$ are ϑ and ϑ^{-1}, where ϑ is given by

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

We choose $h(\lambda)$ to remove the inside root λ. As a result, the equilibrium policy rule can be expressed as

$$h(L) = \left(1-\frac{\vartheta}{\rho}\right)\frac{\lambda(1+\mu)}{\lambda(1+\mu)-\mu\vartheta}\frac{1-\frac{\mu\rho\vartheta}{\lambda(1+\mu)}L}{1-\vartheta L} = \frac{\lambda(1+\mu)}{\lambda(1+\mu)-\mu\vartheta}\left(1-\frac{\mu\rho\vartheta}{\lambda(1+\mu)}L\right)h^{*}(L).$$

A.20 Proof of Proposition 5.1

Based on the single-agent solution in Huo and Pedroni (2020), for any $\alpha \in (-1, 1)$, the aggregate outcome in the beauty-contest game is the same as that in the following pure forecasting problem

$$a_{it} = \mathbb{E}_{it}[\xi_t],$$

with the modification that the precision of the private shocks is discounted by $1 - \alpha$. Therefore, it is sufficient to prove that there does not exist a finite-state representation for the pure forecasting problem.

Suppose that the perceived law of motion of a_t is

$$a_t = \phi(L)\eta_t \equiv \frac{a(L)}{b(L)}\eta_t$$

where $a(L)$ and $b(L)$ are finite polynomials, $b(L)$ does not contain any inside root, and $a(L)$ and $b(L)$ do not contain any common root.

Without loss of generality, assume that $\eta_t \sim \mathcal{N}(0, 1)$, $u_{it} \sim \mathcal{N}(0, \sigma_u^2)$, and $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. Given this perceived law of motion, the signal process can be represented as

$$\boldsymbol{x}_{it} = \mathbf{M}(L)\begin{bmatrix}\widehat{\eta}_t \\ \widehat{u}_{it} \\ \widehat{\varepsilon}_{it}\end{bmatrix}, \quad \text{where} \quad \mathbf{M}(L) = \begin{bmatrix} 1 & \sigma_u & 0 \\ \frac{a(L)}{b(L)} & 0 & \sigma_\varepsilon \end{bmatrix},$$

where $\widehat{\eta}_t, \widehat{\varepsilon}_{it}, \widehat{u}_{it}$ are normalized shocks with unit standard deviation.

Step 1. Factorization. We first construct the spectral factorization of the signal process.

$$\begin{aligned}
&\mathbf{M}(z)\mathbf{M}'(z^{-1}) \\
&= \begin{bmatrix} 1+\sigma_u^2 & \frac{a(z^{-1})}{b(z^{-1})} \\ \frac{a(z)}{b(z)} & \frac{a(z)a(z^{-1})}{b(z)b(z^{-1})} + \sigma_\varepsilon^2 \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 \\ \frac{a(z)}{(1+\sigma_u^2)b(z)} & 1 \end{bmatrix}\begin{bmatrix} 1+\sigma_u^2 & 0 \\ 0 & \frac{a(z)a(z^{-1})}{b(z)b(z^{-1})}\frac{\sigma_u^2}{1+\sigma_u^2} + \sigma_\varepsilon^2 \end{bmatrix}\begin{bmatrix} 1 & \frac{a(z^{-1})}{(1+\sigma_u^2)b(z^{-1})} \\ 0 & 1 \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 \\ \frac{a(z)}{(1+\sigma_u^2)b(z)} & 1 \end{bmatrix}\begin{bmatrix} 1+\sigma_u^2 & 0 \\ 0 & \frac{\sigma_u^2 a(z)a(z^{-1}) + \sigma_\varepsilon^2(1+\sigma_u^2)b(z)b(z^{-1})}{(1+\sigma_u^2)b(z)b(z^{-1})} \end{bmatrix}\begin{bmatrix} 1 & \frac{a(z^{-1})}{(1+\sigma_u^2)b(z^{-1})} \\ 0 & 1 \end{bmatrix}, \\
&= \begin{bmatrix} 1 & 0 \\ \frac{a(z)}{(1+\sigma_u^2)b(z)} & 1 \end{bmatrix}\begin{bmatrix} 1+\sigma_u^2 & 0 \\ 0 & \frac{d^2\lambda(z)\lambda(z^{-1})}{b(z)b(z^{-1})} \end{bmatrix}\begin{bmatrix} 1 & \frac{a(z^{-1})}{(1+\sigma_u^2)b(z^{-1})} \\ 0 & 1 \end{bmatrix},
\end{aligned}$$

where we can choose $\lambda(z)$ so that all the roots of $\lambda(z)$ are outside the unit circle, the roots of $\lambda(z^{-1})$ are inside the unit circle, and d is such that

$$d^2\lambda(z)\lambda(z^{-1}) = \sigma_u^2 a(z)a(z^{-1}) + \sigma_\varepsilon^2(1+\sigma_u^2)b(z)b(z^{-1}).$$

Continuing the factorization,

$$
\begin{aligned}
\mathbf{M}(z)\mathbf{M}'(z^{-1}) &= \begin{bmatrix} 1 & 0 \\ \frac{a(z)}{(1+\sigma_u^2)b(z)} & 1 \end{bmatrix} \begin{bmatrix} \sqrt{1+\sigma_u^2} & 0 \\ 0 & \frac{d\lambda(z)}{b(z)} \end{bmatrix} \begin{bmatrix} \sqrt{1+\sigma_u^2} & 0 \\ 0 & \frac{d\lambda(z^{-1})}{b(z^{-1})} \end{bmatrix} \begin{bmatrix} 1 & \frac{a(z^{-1})}{(1+\sigma_u^2)b(z^{-1})} \\ 0 & 1 \end{bmatrix}, \\
&= \begin{bmatrix} \sqrt{1+\sigma_u^2} & 0 \\ \frac{a(z)}{\sqrt{1+\sigma_u^2}b(z)} & \frac{d\lambda(z)}{b(z)} \end{bmatrix} \begin{bmatrix} \sqrt{1+\sigma_u^2} & \frac{a(z^{-1})}{\sqrt{1+\sigma_u^2}b(z^{-1})} \\ 0 & \frac{d\lambda(z^{-1})}{b(z^{-1})} \end{bmatrix}, \\
&\equiv \mathbf{\Gamma}(z)\mathbf{\Gamma}'(z^{-1}).
\end{aligned}
$$

Since that $b(z)$ does not have inside roots, $\mathbf{\Gamma}(z)$ is analytic inside the unit circle and $\det \mathbf{\Gamma}(z) = d\sqrt{1+\sigma_u^2}\frac{\lambda(z)}{b(z)}$ assures the rank of $\mathbf{\Gamma}(z)$ is 2 everywhere inside the unit circle. Therefore, $\mathbf{\Gamma}(z)$ is a fundamental representation of the signal process $\mathbf{M}(z)$.

Step 2. Inference. In this proof, we consider a more general process for the fundamental than that in the main text. Suppose that ξ_t follows an AR (p) process

$$
\xi_t = \frac{1}{p(L)}\eta_t,
$$

where $p(L)$ does not contain any inside root and its order is p. By the Wiener-Hopf prediction formula, the average optimal forecast about ξ_t is given by

$$
\begin{aligned}
&\overline{\mathbb{E}}_t\left[\frac{1}{p(L)}\eta_t\right] \\
= &\left[\begin{bmatrix} \frac{1}{p(L)} & 0 & 0 \end{bmatrix} \mathbf{M}(L^{-1})\mathbf{\Gamma}(L^{-1})^{-1}\right]_+ \mathbf{\Gamma}(L^{-1})\mathbf{M}(L)\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}\eta_t, \\
= &\left[\begin{bmatrix} \frac{1}{p(L)} & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & \frac{a(L^{-1})}{b(L^{-1})} \\ \sigma_u & 0 \\ 0 & \sigma_\varepsilon \end{bmatrix} \begin{bmatrix} \sqrt{1+\sigma_u^2} & \frac{a(L^{-1})}{\sqrt{1+\sigma_u^2}b(L^{-1})} \\ 0 & \frac{d\lambda(L^{-1})}{b(L^{-1})} \end{bmatrix}^{-1}\right]_+ \begin{bmatrix} \sqrt{1+\sigma_u^2} & 0 \\ \frac{a(L)}{\sqrt{1+\sigma_u^2}b(L)} & \frac{d\lambda(L)}{b(L)} \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \frac{a(L)}{b(L)} \end{bmatrix}\eta_t, \\
= &\begin{bmatrix} \frac{1}{\sqrt{1+\sigma_u^2}p(L)} & \left[\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)}\right]_+ \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{1+\sigma_u^2}} \\ \frac{\sigma_u^2 a(L)}{(1+\sigma_u^2)d\lambda(L)} \end{bmatrix}\eta_t,
\end{aligned}
$$

where the last line follows from the fact that $p(L)$ does not contain any inside root.

Next, we show that the part involving the annihilation operator satisfies

$$
\left[\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)}\right]_+ = \frac{q(L)}{p(L)}, \tag{A.10}
$$

where $q(L)$ is a polynomial in L with the properties that: (1) the order of $q(L)$ is less than or equal to $p-1$; and (2) $q(L)$ does not have any common roots with $p(L)$.[32]

Let n_a and n_λ denote the order of $a(L)$ and $\lambda(L)$, respectively. If $n_a \geq n_\lambda$

$$
\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)} = \frac{\sigma_u^2 a(L^{-1})L^{n_a}}{(1+\sigma_u^2)d\lambda(L^{-1})L^{n_a}p(L)} = \frac{\sigma_u^2 \tilde{a}(L)}{(1+\sigma_u^2)d\tilde{\lambda}(L)L^{n_a-n_\lambda}p(L)},
$$

[32]The second property is not necessary for the proof, but we include it here for completeness.

and otherwise

$$\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)} = \frac{\sigma_u^2 a(L^{-1})L^{n_\lambda}}{(1+\sigma_u^2)d\lambda(L^{-1})L^{n_\lambda}p(L)} = \frac{\sigma_u^2 \tilde{a}(L)L^{n_\lambda-n_a}}{(1+\sigma_u^2)d\tilde{\lambda}(L)p(L)},$$

where $\tilde{a}(L) \equiv a(L^{-1})L^{n_a}$ and $\tilde{\lambda}(L) \equiv \lambda(L^{-1})L^{n_\lambda}$. The order of $\tilde{a}(L)$ is at most n_a. The order of $\tilde{\lambda}(L)$ is n_λ because there needs to be a constant term in $\lambda(L)$. Otherwise, $\lambda(L)$ would contain zero as an inside root, which contradicts the fact that $\lambda(L)$ contains only outside roots. Therefore, in both cases, the order of the denominator of $\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)}$ is greater than that of the numerator.

Denote $\tilde{\lambda}(L) = \prod_{i=1}^{I}(z-\delta_i)^{\tau_i}$ and $p(L) = \prod_{i=1}^{J}(z-b_j)^{\kappa_i}$ where $\sum_{i=1}^{I}\tau_i = n_\lambda$ and $\sum_{i=1}^{J}\kappa_i = p$. By the partial fraction decomposition, there exist constants $\{c_{ik}\}$, $\{d_{jk}\}$, and $\{e_i\}$ such that[33]

$$\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)} = \begin{cases} \sum_{i=1}^{I}\sum_{k=1}^{\tau_i}\frac{c_{ik}}{(z-\delta_i)^k} + \sum_{i=1}^{J}\sum_{k=1}^{\kappa_i}\frac{d_{ik}}{(z-b_i)^k} + \sum_{i=1}^{n_a-n_\lambda}\frac{e_i}{z^i} & \text{if } n_a \geq n_\lambda, \\ \sum_{i=1}^{I}\sum_{k=1}^{\tau_i}\frac{c_{ik}}{(z-\delta_i)^k} + \sum_{i=1}^{J}\sum_{k=1}^{\kappa_i}\frac{d_{ik}}{(z-b_i)^k} & \text{otherwise.} \end{cases} \tag{A.11}$$

By construction, all the roots $\{\delta_i\}$ are inside the unit circle and all the roots $\{b_i\}$ are outside the unit circle. As a result,

$$\left[\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)}\right]_+ = \sum_{j=1}^{J}\sum_{k=1}^{\kappa_j}\frac{d_{jk}}{(z-b_j)^k} = \frac{\sum_{j=1}^{J}\sum_{k=1}^{\kappa_j} d_{jk}(z-b_j)^{\kappa_j-k}\prod_{h\neq j,1\leq h\leq J}(z-b_h)^{\kappa_h}}{\prod_{j=1}^{J}(z-b_j)^{\kappa_j}},$$

where $q(L)$ corresponds to the numerator of the last term. It follows that the order of the numerator is at most $p-1$. In addition, the numerator $q(L)$ cannot contain any common roots with $p(L)$. Otherwise, it implies that $d_{j\kappa_j} = 0$ for some j, which makes the decomposition (A.11) invalid.

Step 3. Verifying Equilibrium Condition. Using condition (A.10), it follows that the implied actual law of motion $\psi(L) = \overline{\mathbb{E}}_t[\xi_t]$ is

$$\psi(L) = \begin{bmatrix} \frac{1}{\sqrt{1+\sigma_u^2}p(L)} & \left[\frac{\sigma_u^2 a(L^{-1})}{(1+\sigma_u^2)d\lambda(L^{-1})p(L)}\right]_+ \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{1+\sigma_u^2}} \\ \frac{\sigma_u^2 a(L)}{(1+\sigma_u^2)d\lambda(L)} \end{bmatrix} = \frac{d\lambda(L)+\sigma_u^2 q(L)a(L)}{d(1+\sigma_u^2)p(L)\lambda(L)}.$$

In equilibrium, the perceived law of motion and the actual law of motion of the aggregate outcome need to be consistent with each other.

First, we consider the case that $a(L)$ is not a constant and has at least one root. To match with $\phi(L)$, the numerator of $\psi(L)$ needs to have all the roots of $a(L)$, so $\lambda(L)$ in the numerator can be decomposed as

$$\psi(L) = \frac{(d\lambda_1(L)+\sigma_u^2 q(L))a(L)}{d(1+\sigma_u^2)p(L)\lambda(L)},$$

where $\lambda_1(L)$ is a polynomial such that $\lambda(L) = \lambda_1(L)a(L)$. On the other hand, together with the definition of $\lambda(L)$, $\lambda(L) = \lambda_1(L)a(L)$ implies that

$$d^2\lambda(L)\lambda(L^{-1}) = d^2\lambda_1(L)a(L)\lambda_1(L^{-1})a(L^{-1}) = \sigma_u^2 a(L)a(L^{-1}) + \sigma_\varepsilon^2(1+\sigma_u^2)b(L)b(L^{-1}).$$

It follows that $b(L)b(L^{-1})$ can be expressed as $\sigma_\varepsilon^2(1+\sigma_u^2)b(L)b(L^{-1}) = a(L)a(L^{-1})b_1(L)b_1(L^{-1})$, where $b_1(L)$ is a polynomial.

[33]These coefficients can be found by applying the Heaviside theorem.

That is, $b(L)b(L^{-1})$ has to contain all the roots of $a(L)a(L^{-1})$. This is possible only if all the roots of $a(L)$ are reciprocals of the roots of $b(L)$ as $a(L)$ and $b(L)$ do not contain any common root. Then, all the roots of $a(L)$ must be inside the unit circle given that all the roots of $b(L)$ are outside the unit circle. It follows that $\lambda(L)$ cannot have any common roots with $a(L)$ as all the roots of $\lambda(L)$ are outside the unit circle by construction. This is a contradiction. Therefore, the numerator of $\psi(L)$ cannot have all of the roots of $a(L)$ to match with $\phi(L)$.

Second, we consider the case where $a(L) = \chi$ is simply a constant. If so, $\psi(L)$ becomes

$$\psi(L) \quad = \quad \frac{d\lambda(L) + \sigma_u^2 q(L)\chi}{d(1+\sigma_u^2)p(L)\lambda(L)},$$

and by definition, $\lambda(L)$ satisfies

$$d^2\lambda(L)\lambda(L^{-1}) = \sigma_u^2\chi^2 + \sigma_\varepsilon^2(1+\sigma_u^2)b(L)b(L^{-1}). \tag{A.12}$$

If $b(L)$ is constant, then both $\psi(L)$ and $\lambda(L)$ must be constants. But this leads to a contradiction since the order of $q(L)$ is at most $p-1$ and the order of the denominator is p.

If $b(L)$ have at least one root, condition (A.12) implies that the order of $\lambda(L)$ is the same as the order of $b(L)$, and $\lambda(L)$ and $b(L)$ cannot have any common roots. Then, to match with $\phi(L)$, the denominator of $\psi(L)$ must not have any root of $\lambda(L)$, which requires that the roots of the numerator $d\lambda(L) + \sigma_u^2 q(L)\chi$ contain all the roots of $\lambda(L)$. Consequently,

$$\psi(L) = \frac{\lambda(L)[d + q_1(L)]}{d(1+\sigma_u^2)p(L)\lambda(L)} = \frac{d + q_1(L)}{d(1+\sigma_u^2)p(L)},$$

where $q_1(L)$ is such that $q(L)\sigma_u^2\chi = \lambda(L)q_1(L)$.

Let n_b be the order of $b(L)$, which is the order of $\lambda(L)$ as well. Since the order of $q(L)$ is at most $p-1$, the order of $d+q_1(L)$ is at most $p-n_b-1$. Meanwhile, the order of the denominator is simply p. The difference of the orders between the denominator and the numerator is then at least n_b+1. This cannot equal to the difference between the order of $b(L)$ and the order of $a(L)$, which is n_b. This is a contradiction again and $\psi(L)$ cannot match with $\phi(L)$.