Rational Expectations Models with Higher Order Beliefs*

Zhen Huo — Yale University

Naoki Takayama — Cabinet Office, Government of Japan

Monday 12th February, 2018

Abstract

This paper develops a general method of solving rational expectations models with higher order beliefs. Higher order beliefs are crucial in an environment with dispersed information and strategic complementarity, and the equilibrium policy depends on infinite higher order beliefs. It is generally believed that solving this type of equilibrium policy requires an infinite number of state variables (Townsend, 1983). This paper proves that the equilibrium policy rule can always be represented by a finite number of state variables if the signals observed by agents follow an ARMA process, in which case we obtain a general solution formula. We also prove that when the signals contain endogenous variables, a finite-state-variable representation of the equilibrium may not exist. The key innovation in our method is to use the factorization identity and Wiener filter to solve signal extraction problems conditional on infinite signals. This method can be used in a wide range of applications.

Keywords: Higher order beliefs, Infinite regress problem, Wiener filter.

*Huo: zhen.huo@yale.edu. Takayama: naoki.takayama@gmail.com. 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 especially grateful for comments from Anmol Bhandari, V.V. Chari, Patrick Kehoe, Fabrizio Perri, Thomas Sargent, and Kei-Mu Yi. We also thank Sushant Acharya, Manuel Amador, Yan Bai, Luigi Bocola, Katarina Borovickova, Jaroslav Borovicka, Fend Dong, Si Guo, Boyan Jovanovic, Kyle Herkenhoff, Guido Lorenzoni, Shihui Ma, Virgiliu Midrigan, Kris Nimark, Alessandro Peri, Edouard Schaal, Kjetil Storesletten, Gianluca Violante, Chunying Xie, and the participants of the Labor Workshop at the University of Minnesota, Federal Reserve Bank of Minneapolis, and New York University for helpful comments. Zhen Huo thanks the University of Minnesota Doctoral Dissertation Fellowship for financial support.

1 Introduction

In many economic models with information frictions, an agent's payoff depends on her own actions, the actions of others, and some unknown economic fundamentals. Rational behaviors not only depend on an agent's beliefs on economic fundamentals, but also depend on higher order beliefs, that is, agents' beliefs of others' beliefs, agents' beliefs of others' beliefs of others' beliefs, and so on. If the economic fundamentals are persistent over time and hence the past information is worth keeping track of, forecasting all the higher order beliefs would require an infinite number of priors of them, which would amount to an infinite number of state variables. This type of problem is known as the *infinite regress problem*, and has been explored by a large number of works.[1]

The difficulty of solving models with higher order beliefs lies in the fact that inferring others' action requires the functional form of the policy rule in the first place, but the policy rule is the solution to the inference problem. As argued in Townsend (1983), if an agent assumes that other agents keep track of n state variables, he in turn needs to keep track of $n+1$ state variables (the prior of the economic fundamental and the n priors of others' state variables). Therefore, the equilibrium policy rule does not permit a finite-state representation. In terms of higher order beliefs, to predict $k-$th order belief requires at least k state variables, and to predict all the higher order beliefs requires infinite state variables. In light of these considerations, it is generally believed that an infinite number of state variables are needed to solve this type of model.

In this paper, we pursue the following question. With higher order beliefs, is it really impossible to find a small set of state variables that are sufficient statistics for agents to make the optimal inference? If possible, how do we find these state variables and what are the laws of motion for these variables? If it does require an infinite number of state variables, how do we approximate the true solution with a finite number of state variables?

Our first main result is that given a linear rational expectations model, when observed signals follow ARMA processes, the equilibrium policy rule always allows a finite-state representation. In applications with exogenous information, the ARMA signal structure is a common specification. Like in standard problems with common information, solving for the equilibrium

[1] A partial list of these works includes Chari (1979), Townsend (1983), Singleton (1987), Sargent (1991), Kasa (2000), Woodford (2003), Lorenzoni (2009), Angeletos and La'O (2010), Hellwig and Venkateswaran (2009), Rondina and Walker (2017), and so on.

requires finding the fixed point in a functional space. Unlike in standard models, when higher order beliefs are involved, it is difficult to determine the sufficient state variables in the first place. Given this difficulty, we start from the state space that is spanned by the entire history of signals. This implies that solving for the equilibrium requires solving for a lag polynomial with an infinite number of coefficients. Our work is based on Whiteman (1983) and Kasa (2000). The idea is to transform the problem which solves for a lag polynomial into a simpler problem which solves for an analytic function. When signals follow an ARMA process, we prove that the equilibrium policy rule, the lag polynomial, is also of the ARMA form. Therefore, there exists finite-state representation for the equilibrium policy rule. The intuition for the finite-state representation is that agents do not directly care about each of the higher order beliefs, but they only care about a specific linear combination of all the higher order beliefs. The latter requires less information and admits a finite-state representation.

We extend the work of Kasa (2000) and others in two important ways. First, we do not restrict the number of signals to being equal to the number of shocks. A necessary step in the inference problem with infinite sample is to find the Wold (fundamental) representation for the signal process. Previous works rely on the Blaschke matrices to find the fundamental representation, which require that the number of signals equals the number of shocks.[2] We adopt a different approach for finding the Wold representation. We show that one can first convert the signal process into its state-space, and then use the innovation representation and factorization identity to solve for the Wold representation conveniently. This procedure works for any information structure that follows an ARMA process: it is not restricted by the number of signals or the number of shocks. In general signal extraction problems, there are more shocks than signals, as discussed in Nimark (2017). This restriction that there has to be the same number of signals as shocks is indeed violated in many applications, such as Woodford (2003), Angeletos and La'O (2010) and Angeletos and La'O (2013). When this restriction is actually satisfied, agents often learn 'too much', in the sense that the prediction error is not long-lasting, because there are insufficient numbers of noisy shocks to really confuse them, unless assuming a confounding shock process in the first place.[3] In both Kasa (2000) and Acharya (2013), agents can learn the true state of the economy after one period. When there are more shocks than signals, agents never fully learn the true state of the economy and the prediction error is typically persistent. As a result, the model economy features more

[2]See Rondina and Walker (2017), Kasa, Walker, and Whiteman (2014) and Acharya (2013) for example. Walker (2007) solved a special signal process with more shocks than signals.

[3]In Rondina and Walker (2017), they assume a non-invertible shock process.

relevant and richer dynamics.

Secondly, we allow agents to solve a general signal extraction problem. The majority of existing literature that applies the frequency-domain technique only studies a pure forecasting problem. That is, only future values of signals are pay-off relevant. To forecast future signals, one can simply use the Hansen-Sargent formula. In the examples presented in this paper, agents need to solve a generic signal extraction problem conditional on infinite observables. The Hansen-Sargent formula does not apply in these environments. Instead, we apply the Wiener-Hopf prediction formula, which is well suited for these types of problems and includes Hansen-Sargent formula as a special case. Applying the Wiener-Hopf prediction formula in the univariate case has been discussed in Sargent (1987) and applied in Rondina (2008). In this paper, we extend the application to multivariate case, and establish a number of its general properties under ARMA signal structure.

We illustrate our method in various applications where economic fundamentals or noises are persistent. In the first application, we explore the classcial static beauty contest model as in Morris and Shin (2002), where asymmetric information and strategic complementarity make higher order beliefs relevant. We obtain a sharp analytical solution, agents to show the inertia induced by higher order uncertainty, which was analyzed numerically by Woodford (2003) and Angeletos and La'O (2010). In the second application, we compare the static beauty contest model with dynamic beauty contest models, in which the best reponse features forward-looking or backward-looking behavior. These models are common when agents need to make intratemporal decisions, and we show how the the persistence of agents' action depends on the horizon of their strategic motive. In the third application, we extend the model in Angeletos and La'O (2013) with persistent shocks. Different from the last two applications, each agent cares about their idiosyncratic fundamental and interacts with a random agent every period. These features complicate the inference problem, but our method delivers the result and can be compared with the solution based on the heterogeneous-prior assumption in Angeletos and La'O (2013). In the first three applications, agents share the same best response and the equilibrium is symmetric. In the last application, we consider the case where agents differ in their best responses, and therefore have different policy rules. Even thought these applications provide a number of novel results, they have not taken full advantage of our method. In Huo and Takayama (2017), we show that the method can be applied to more complicated quantitative models.

The result on finite-state representation can be applied for cases where agents solve their inference problem given an exogenous ARMA signal process. We also explore cases when agents observe signals that contain information which is endogenously determined in the equilibrium. We label them as problems with endogenous information. The equilibrium with endogenous information imposes an additional cross-equation restriction, in the sense that the perceived law of motion has to be consistent with the realized law of motion. The endogenous variable that appears in the signal has an information role as well.

Our second main result is that we provide examples with endogenous information, in which the equilibrium cannot be represented by finite state variables.[4] In the model, agents can observe the past aggregate action with noises, and the aggregate action actually follows an infinite order process. This result is somewhat surprising given that the exogenous driving force of the economy is very simple. It should be clear that it is not because of the infinite higher order beleifs that agents have to keep track of infinite state variables. For each individual, they still take the signal process as exogenously given, even though the signals contain an equilibrium object. From our first main result, once the endogenous variable follows a finite ARMA process, the individual policy rule has a finite-state representation. If the endogenous variable does not follow a finite ARMA process, the signal received by agents cannot follow a finite ARMA process. Note that in Kasa (2000) and other papers where the number of signals is the same as the number of shocks, the equilibrium permits a finite-state representation even with endogenous information. When we allow for a more general information process, this result does not hold any more.

This finding is interesting from a theoretical point of view, but it also implies that finding the exact solution is no longer possible. To solve the problem with endogenous information, we approximate the law of motion of the endogenous variable that shows up in signals by an ARMA process. Note that this ARMA approximation method is different from Sargent (1991) and others in an important way. Even though we approximate the law of motion of the endogenous variable, each individual still faces the infinite regress problem. Using our method, each individual's policy rule is solved exactly.

Related literature Our paper is closely related to the literature that attempts to solve the infinite regress problem. Broadly speaking, there are two approaches to solving the infinite regress problem. The first approach is to short-circuit the infinite regress problem by modifying the original problems. For example, by assuming that information becomes public after certain

[4] Chari (1979) proved a similar impossibility theory.

periods, the relevant state space is finite and one can use the Kalman filter. A partial list of literature that employs this method includes Townsend (1983), Hellwig and Venkateswaran (2009), Lorenzoni (2009), Bacchetta and Wincoop (2006), Perez and Drenik (2015). This assumption is unsatisfying from a modeling perspective, and it is proved by Walker (2007), Kasa (2000) and Pearlman and Sargent (2005) that the approximate solution can be very different from the true solution. Another type of approximation is developed by Nimark (2008) and Nimark (2017). The idea is that quantitatively only a finite order of higher order beliefs matter for the equilibrium, based on the observation that the effects of higher order beliefs diminish as the order increases. This method provides important insights into the nature of the higher order beliefs, but this method can be difficult to implement when the degree of strategic complementarity is strong, or when the model is complicated to express the policy rule in terms of higher order beliefs. Sargent (1991) approximated the equilibrium via the ARMA process. The forecasting problem is transformed into fitting vector ARMA models, which is particularly useful when agents do not need to solve a pure forecasting problem.

The second approach is to solve the infinite regress problem exactly without approximation. Kasa (2000) first uses the frequency-domain method to solve the Townsend (1983) original problem and found that agents actually share the same belief and there is no infinite regress problem. Walker (2007), Rondina and Walker (2017), and Kasa, Walker, and Whiteman (2014) apply the frequency-domain method to study various asset pricing models proposed by Futia (1981) and Singleton (1987). Acharya (2013) applies this method to study the effects of noises on business cycles. These papers assume that the number of shocks equals the number of signals in order to obtain a closed form solution when there exists endogenous information. Huo and Pedroni (2017) obtains a simple solution to static beauty contest models with general information structure. Our paper complements this line of literature. For exogenous information, a much broader class of models can be solved by our method.

Our applications in this paper complement the literature on macroeconomics with higher order beliefs. We obtain analytical solutions for models closely related to Woodford (2003), Angeletos and La'O (2010), Nimark (2017), and Angeletos and La'O (2013). We believe our method is also useful in solving models similar to Lorenzoni (2009), Hellwig and Venkateswaran (2009), Graham and Wright (2010) and others. In our companion paper (Huo and Takayama, 2017), we study a business cycle model driven by confidence shocks. We characterize how information frictions affect the persistence and variance of output, and show that the confidence shock could be an important factor in explaining business cycles.

The rest of the paper is organized as follows. Section 2 sets up a two-player model to introduce higher order beliefs and the infinite regress problem. Section 3 presents the main results. We first show how to jointly use the Kalman filter and the Wiener-Hopf prediction formula to form the optimal expectation with entire history of signals. This enable us to obtain a finite-state representation for a rational expectations model with higher order beliefs. Section 4 presents several applications with persistent information where analytical solutions are available. Section 5 explores the case in which the signals contain an endogenous variable. We prove that the equilibrium policy rule does not have a finite-state representation in this environment. Section 6 concludes.

2 A Two-Player Model

In this section, we present a simple two-player model with the infinite regress problem. This model naturally assigns an important role to infinite higher order beliefs, and numerous variations of it have been used in the literature. We use this model to define some concepts that will become useful in Section 3.

2.1 Model setup

Consider a game between two agents i and j. Time is discrete and lasts forever. In period t, agents' payoff depends on a common persistent economic fundamental ξ_t. In addition, the payoff features strategic complementarity or substitutability, and it also depends on the action of the other agent. Information frictions prevent agents from perfectly observing ξ_t or the action of the other agent.

We assume that the best response of agent i, denoted by y_{it}, has to satisfy

$$y_{it} = \mathbb{E}[\xi_t|\Omega_{it}] + \alpha\mathbb{E}[y_{jt}|\Omega_{it}], \tag{2.1}$$

where $\alpha \in (0, 1)$ determines the strength of strategic complementarity and Ω_{it} denotes the information set of agent i at time t. We consider a symmetric scenario in which agent j uses the same best response and has the same information structure as agent i. Note that agents make a purely static decision every period, and the link across different periods is only through the information set. There are various micro-foundations that lead to this specification, such as Woodford (2003) and Angeletos and La'O (2010). For now we only focus on this abstract form and discuss its general properties. The information structure of the model is specified as follows.

Signal process We assume that ξ_t follows a covariance stationary ARMA (p, q) process

$$\xi_t = \sum_{k=1}^{p} \rho_k \xi_{t-k} + \sum_{k=0}^{q} \theta_k \eta_{t-k},$$

where $\eta_t \sim \mathcal{N}(0, \sigma_\eta)$. As opposed to observing ξ_t directly, agents receive n signals that are related to ξ_t. These signals are simply the sum of ξ_t and some idiosyncratic noises

$$\mathbf{x}_{it} \equiv \begin{bmatrix} x_{it}^1 \\ \vdots \\ x_{it}^n \end{bmatrix} = \begin{bmatrix} \xi_t + \epsilon_{it}^1 \\ \vdots \\ \xi_t + \epsilon_{it}^n \end{bmatrix}, \tag{2.2}$$

where $\epsilon_{it}^\tau \sim \mathcal{N}(0, \sigma_\tau^2)$ for $\tau \in \{1, \ldots, n\}$. The information set of agent i at time t contains all the signals he has received up to time t

$$\Omega_{it} = \left\{ \mathbf{x}_{it}, \mathbf{x}_{it-1}, \mathbf{x}_{it-2}, \ldots \right\}. \tag{2.3}$$

Agent j receives signals of ξ_t that are corrupted by her idiosyncratic noises ϵ_{jt}^τ. As a result, these two agents do not share the same information set. To simplify notation, we will use $\mathbb{E}_{it}[\,\cdot\,]$ to denote the conditional expectation operator $\mathbb{E}[\,\cdot \mid \Omega_{it}]$ from now on.

The information structure and the best response we have specified above are very special. We will relax these assumptions in the next section.

2.2 Higher order beliefs

The best response of agent i is given by equation (2.1), and the same rule applies to agent j,

$$y_{jt} = \mathbb{E}_{jt}[\xi_t] + \alpha \mathbb{E}_{jt}[y_{it}]. \tag{2.4}$$

We can repeatedly substitute equation (2.4) into equation (2.1), and vice versa, which leads to

$$\begin{aligned} y_{it} &= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_{jt}] \\ &= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\left[\mathbb{E}_{jt}[\xi_t] + \alpha \mathbb{E}_{jt}[y_{it}]\right] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\mathbb{E}_{jt}[\xi_t] + \alpha^2 \mathbb{E}_{it}\mathbb{E}_{jt}[y_{it}] \\
&= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\mathbb{E}_{jt}[\xi_t] + \alpha^2 \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}[\xi_t] + \alpha^3 \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}[y_{jt}] \\
&\;\vdots \\
&= \sum_{k=0}^{\infty} \alpha^k \, \mathbb{E}_{it}^{k+1}[\xi_t],
\end{aligned} \tag{2.5}
$$

where $\mathbb{E}_{it}^{k}[\xi_t]$ stands for k-th order belief. These higher order beliefs are defined recursively as follows

$$
\begin{aligned}
\mathbb{E}_{it}^{1}[\xi_t] &= \mathbb{E}_{it}[\xi_t] \\
\mathbb{E}_{it}^{2}[\xi_t] &= \mathbb{E}_{it}\mathbb{E}_{jt}[\xi_t] \\
\mathbb{E}_{it}^{k}[\xi_t] &= \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}^{k-2}[\xi_t], \text{ for } k = 3, 5, 7, \ldots \\
\mathbb{E}_{it}^{k}[\xi_t] &= \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}^{k-2}[\xi_t], \text{ for } k = 4, 6, 8, \ldots
\end{aligned}
$$

Crucially, agents have heterogeneous information sets, and the law of iterated expectations does not apply. Hence, the optimal action y_{it} depends on all the higher order beliefs. Mathematically, the means of all these higher order beliefs can be calculated by the standard Kalman filter, but there are an infinite number of such objects to be calculated. One may think that if a certain pattern of these higher order beliefs is found, these beliefs may be summarized in a compact way. This approach works for a static system, such as in Morris and Shin (2002). However, it does not work in general, due to a growing complexity with the order of beliefs. Forecasting all of these higher order beliefs requires an infinite number of priors of these beliefs, and these priors are functions of the entire history of agents' signals. As a result, it is generally believed that the policy rule has to include the entire history of signals as state variables.

It should be clear that higher order beliefs is a particular representation of agents' optimal action. Agents do not have to use all higher order beliefs to determine their optimal action. If the laws of motion of other agents' actions are known, then a first order belief about the fundamental and others' actions will be sufficient. In fact, this is the approach we will take in this paper. Still, higher order beliefs are very helpful in understanding some properties of the effects of dispersed information.

2.3 Equilibrium

By the higher order belief representation of the best response, it is clear that the optimal policy is linear in the signals as it is ultimately a result of optimal forecasting. The linear policy rule of agent i belongs to the space spanned by square-summable linear combinations of current and past realizations of $\mathbf{x}_{it}$. We use $\mathcal{H}_t^x$ to denote this space. The optimal action is a linear combination of current and past signals

$$y_{it} = \sum_{k=0}^{\infty} h_{1k} x_{it-k}^1 + \sum_{k=0}^{\infty} h_{2k} x_{it-k}^2 + \ldots \sum_{k=0}^{\infty} h_{nk} x_{it-k}^n, \tag{2.6}$$

and it is obvious that $y_{it} \in \mathcal{H}_t^x$. In standard models without higher order beliefs, the policy rule still depends on the entire history of signals, but a finite number of state variables can be easily found to effectively summarize the past information. In contrast, due to the infinite higher order beliefs, it is very difficult to figure out whether there exists a finite number of state variables in the first place (even though later on we prove that this is indeed the case) and the laws of motion of these state variables given they exist. In principle, agents have to keep track of the entire history of signals.

To fix the language, we fist define the casual and noncasual stationary lag polynomials.

Definition 2.1. *$\phi(L)$ is a non-casual stationary lag polynomial if*

$$\phi(L) = \sum_{k=-\infty}^{\infty} \phi_k L^k, \tag{2.7}$$

and $\{\phi_k\}_{k=-\infty}^{\infty} \in \ell^2$. $\phi(L)$ is a casual-stationary lag polynomial if it does not contain L with negative powers

$$\phi(L) = \sum_{k=0}^{\infty} \phi_k L^k. \tag{2.8}$$

More compactly, we can use lag polynomials to describe the policy rule

$$y_{it} = h_1(L) x_{it}^1 + h_2(L) x_{it}^2 + \ldots + h_n(L) x_{it}^n = \mathbf{h}(L)\mathbf{x}_{it}. \tag{2.9}$$

In equilibrium, agents cannot use signals realized in the future, and hence the policy rule $\mathbf{h}(L)$ should be casual-stationary lag polynomials. The definition of the equilibrium is straightforward.

Definition 2.2. *Given the signal process (2.2), the equilibrium of model (2.1) is a causal-stationary policy rule* $\mathbf{h}(L) = \begin{bmatrix} h_1(L) & h_2(L) & \dots & h_n(L) \end{bmatrix}$*, such that*

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha\, \mathbb{E}_{it}[y_{jt}],$$

where

$$y_{it} = \mathbf{h}(L)\mathbf{x}_{it}, \quad y_{jt} = \mathbf{h}(L)\mathbf{x}_{jt}.$$

To solve for the equilibrium policy rules $\mathbf{h}(L)$, the difficulty lies in how to solve the inference problem

$$\mathbb{E}_{it}\left[y_{jt}\right] = \mathbb{E}_{it}\left[\mathbf{h}'(L)\mathbf{x}_{jt}\right],$$

in which the variable to be predicted is with infinite states. The Kalman filter requires the predicted variable to have finite states, and therefore it is inapplicable for this type of the problem. In contrast, the Wiener filter can solve the inference problem that is conditional on infinite observables, and it allows the predicted variable to have infinite states (the details of these two filters are discussed in the next section). The next section will show, after solving $\mathbb{E}_{it}[y_{jt}]$, it turns out that $\mathbf{h}(L)$ are of finite ARMA type, and it allows a finite-state representation.

3 Methodology: General Linear Rational Expectations Models

In this section, we develop the method that solves the general rational expectations models with higher order beliefs. We first lay out the structure of the model and the signal process. We then show how to prove our main results in steps. In this section, we focus on symmetric equilibrium in which all agents share the same information an payoff structure, and therefore use the same policy rule. However, the method and the finite-state representation result extends to models where agents have heterogeneous payoff or information structures (see subsection 4.4 for an example).

3.1 Rational expectations models

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

Signal process Assume that the signals observed by an individual agent follows

$$\mathbf{x}_t = \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^n \end{bmatrix} = \begin{bmatrix} M_{11}(L) & \dots & M_{1m}(L) \\ \vdots & \ddots & \vdots \\ M_{n1}(L) & \dots & M_{nm}(L) \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} = \mathbf{M}(L)\mathbf{s}_t, \tag{3.1}$$

where the signal $\mathbf{x}_t$ is a stochastic $n \times 1$ vector and the shock $\mathbf{s}_t$ is a stochastic $m \times 1$ vector. We normalize the co-variance matrix of $\mathbf{s}_t$ to be an identity matrix. The natural requirement is that $M_{ij}(L)$ is causal stationary. The information set is $\Omega_t = \mathbf{x}^t = \{\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{x}_{t-2}, \dots\}$.

Choice variable We assume that each individual agent has r choice variables, which are functions of their signals:

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{rt} \end{bmatrix} = \begin{bmatrix} h_{11}(L) & \dots & h_{1n}(L) \\ \vdots & \dots & \vdots \\ h_{r1}(L) & \dots & h_{rn}(L) \end{bmatrix} \begin{bmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{bmatrix} = \mathbf{h}(L)\mathbf{x}_t. \tag{3.2}$$

Here, $\mathbf{h}(L)$ is the matrix of equilibrium policy rules we need to solve. We only require that each element in $\mathbf{h}(L)$ is causal stationary, but do not impose that $\mathbf{h}(L)$ admits a finite ARMA representation in the first place (even though we prove this is indeed the case later).

Endogenous variables related to other agents' actions The optimal policy may depend on other agents' actions or depend on some aggregate endogenous variables. With information frictions, these variables may not be observed perfectly, but matter for agents' payoff. Let $\widetilde{\mathbf{y}}_t$ denote the vector of endogenous variables which are chosen by other agents. In this section, we focus on symmetric equilibrium in which all agents use the same policy rule. They are related to the choice variable $\mathbf{y}_t$ in the following way

$$\widetilde{\mathbf{y}}_t = \mathbf{h}(L)\mathbf{M}(L)\mathbf{\Lambda}\mathbf{s}_t \tag{3.3}$$

If $\mathbf{\Lambda}$ is an identity matrix, then others' action will be the same as the action of the particular individual, $\mathbf{y}_t$. However, the actions of others may only depend on a subset of $\mathbf{s}_t$ and depend on shocks other than $\mathbf{s}_t$. The matrix $\mathbf{\Lambda}$ selects the shocks that also affect other agents, which are common shocks in the economy. For the shocks that are uncorrelated with $\mathbf{s}_t$, the best forecasts of those shocks conditional on $\{\mathbf{x}_t\}$ are simply zero.

Model Assume the optimal action $\mathbf{y}_t$ needs to satisfy the following linear system in equilibrium

$$\mathbb{E}\left[\mathbf{P}(L)\mathbf{y}_t + \mathbf{Q}(L)\widetilde{\mathbf{y}}_t + \mathbf{R}(L)\mathbf{s}_t \middle| \mathbf{x}^t\right] = 0. \tag{3.4}$$

These matrices $\mathbf{P}(L)_{r\times r}, \mathbf{Q}(L)_{r\times r}$ and $\mathbf{R}(L)_{r\times m}$ depend on structural parameters that result from optimal conditions and resource constraints, and they are not related to agents' endogenous choices. These matrices are allowed to be non-casual stationary. This system of equations incorporates the possibilities that the choice variables $\mathbf{y}_t$ depend on the past, the current and the future values of the endogenous variables of others, the exogenous variables, and her own actions. This specification includes the majority of applications that one may encounter.

Special case with Common Information Model (3.4) includes two special cases in which all agents share the same information and there is no need to infer others' action. Even though agents share the same information, they may not necessarily observe the underlying states.

1. Perfect information.

$$\mathbb{E}\left[\mathbf{P}(L)\mathbf{y}_t + \mathbf{Q}(L)\mathbf{y}_t + \mathbf{R}(L)\mathbf{s}_t \middle| \mathbf{s}^t\right] = 0. \tag{3.5}$$

   In standard real business cycle models and New Keynesian models without information frictions, the underlying shocks $\{\mathbf{s}_t\}$ are observed directly by agents. That is, the space spanned by shocks is the same as the space spanned by signals, $\mathcal{H}_t^s = \mathcal{H}_t^x$. Meanwhile, because all the shocks are observed directly, the actions of other agents are also known perfectly. As a result, the expectations can be calculated easily.

2. Imperfect information, but no need to compute higher order beliefs [5]

$$\mathbb{E}\left[\mathbf{P}(L)\mathbf{y}_t + \mathbf{Q}(L)\mathbf{y}_t + \mathbf{R}(L)\mathbf{s}_t \middle| \mathbf{x}^t\right] = 0. \tag{3.6}$$

   This is the case in which information frictions still exist, i.e., $\mathcal{H}_t^x \subset \mathcal{H}_t^s$, but there is no need to infer others' choices. Agents only need to infer the exogenous variables $\mathbf{R}(L)\mathbf{s}_t$, and standard Kalman filter will be sufficient in solving the problem.

In both of these two cases, all agents' signals are driven by the same shocks, and therefore, $\mathbf{\Lambda} = \mathbf{I}$. This property is important in determining the persistence of the policy rule which will be discussed in subsection 3.4.

The solution to model (3.4) is defined as follows

Definition 3.1. *Given the signal process (3.1), an equilibrium is a matrix of causal stationary lag polynomials* $\mathbf{h}(L)$ *such that*

1. *The policy rule* $\mathbf{h}(L)$ *solves*

$$\mathbb{E}\left[\mathbf{P}(L)\mathbf{h}(L)\mathbf{x}_t + \mathbf{Q}(L)\widetilde{\mathbf{y}}_t + \mathbf{R}(L)\mathbf{s}_t \middle| \mathbf{x}^t\right] = 0$$

2. *Others' action* $\widetilde{\mathbf{y}}_t$ *is consistent with* $\mathbf{h}(L)$

$$\widetilde{\mathbf{y}}_t = \mathbf{h}(L)\mathbf{M}(L)\mathbf{\Lambda}\mathbf{s}_t$$

In this section, we aim to answer the following questions:

1. Under what conditions does a unique solution to this model exist?
2. Supposing there indeed exists a solution $\mathbf{h}(L)$ that solves the model, how to find the formula of $\mathbf{h}(L)$?
3. Does the solution admit a finite-state representation that allows agents to summarize the past information using a small set of sufficient statistics?

[5] This case is also discussed in Baxter, Graham, and Wright (2011).

To answer these questions, we further make the following assumptions.

Assumption 3.1. *The signal* $\mathbf{x}_t$ *follows a finite ARMA process, i.e., each element of* $\mathbf{M}(L)$ *is a rational function of the lag operator* L

$$M_{ij}(L) = \frac{a_{ij}(L)}{b_{ij}(L)} = \frac{\sum_{k=0}^{q_{ij}} \alpha_{ijk} L^k}{\sum_{k=0}^{p_{ij}} \delta_{ijk} L^k}, \tag{3.7}$$

and all the roots of $b_{ij}(L)$ *is within the unit circle.*

Assumption 3.2. *The elements in matrices* $\mathbf{P}(L)$, $\mathbf{Q}(L)$, *and* $\mathbf{R}(L)$ *are rational functions of the lag operator* L. *Furthermore,*

$$\mathbf{P}(L) = \frac{\widehat{\mathbf{P}}(L)}{\prod_{k=1}^{d}(L-\beta_k)}, \quad \mathbf{Q}(L) = \frac{\widehat{\mathbf{Q}}(L)}{\prod_{k=1}^{d}(L-\beta_k)}, \quad \mathbf{R}(L) = \frac{\widehat{\mathbf{R}}(L)}{\prod_{k=1}^{d}(L-\beta_k)}, \tag{3.8}$$

where $|\beta_k| < 1$ *and the expansions of* $\widehat{\mathbf{P}}(L), \widehat{\mathbf{Q}}(L), \widehat{\mathbf{R}}(L)$ *are casual lag polynomials.*

Assumption 3.1 and 3.2 essentially assume that $\mathbf{M}(L)$, $\mathbf{P}(L)$, $\mathbf{Q}(L)$, and $\mathbf{R}(L)$ are all of a finite ARMA structure. The potential inside poles $\{\beta_k\}$ capture the possibility that agents' best response depend on their own future actions, others' future actions, or future fundamentals.[6] Under these assumptions, Theorem 1 answers the first two questions, and Proposition 3.3 gives a positive answer to the third question. In short, these results imply that higher order beliefs do not necessarily create infinite state variables. It is always possible to use a small set of variables to summarize past information, given that the signals follow finite ARMA processes.

The proofs of our main results are quite lengthy and involve a number of building blocks. The main steps that lead to Theorem 1 and Proposition 3.3 are listed below

Step 1: Under Assumption 3.1, find the state-space and Wold representation of the signal process.

Step 2: Use Wiener filter to solve the inference problem in model (3.4).

Step 3: Transform the infinite-dimension problem of solving the sequences of coefficients in the lag polynomials into the finite-dimension problem of solving a system of analytic functions.

[6] Note that it is not necessary the case that $\mathbf{P}(L)$, $\mathbf{Q}(L)$, and $\mathbf{R}(L)$ share the same inside poles. One can always set $\widehat{\mathbf{P}}(L)$, $\widehat{\mathbf{Q}}(L)$, and $\widehat{\mathbf{R}}(L)$ to remove the poles.

Step 4: Find the finite-state representation of the equilibrium policy rule.

3.2 State-space representation, Factorization Identity, and Wold representation

We need the Wold representation of the signal process for the following reason. All the prediction is conditional on the observed signals, but ultimately, the linear projection is on the space spanned by shocks. The original underlying shocks $\{\mathbf{s}^t\}$ contain more information than the signals $\{\mathbf{x}^t\}$, and the prediction conditional on $\{\mathbf{s}^t\}$ is different from the prediction conditional on $\{\mathbf{x}^t\}$. The Wold representation provides a new sequence of shocks $\{\mathbf{w}^t\}$. Different from the underlying shocks $\{\mathbf{s}^t\}$, the space spanned by the signals $\{\mathbf{x}^t\}$ is equivalent to the space spanned by $\{\mathbf{w}^t\}$, and we can conduct the linear projection on $\{\mathbf{w}^t\}$. Given a finite ARMA signal process, we present how to find its state-space representation and Wold representation using the factorization identity.

Lemma 3.1. *Under Assumption 3.1, the signal process admits at least one state-space representation, in which the state equation is*

$$\mathbf{Z}_t = \mathbf{F}\mathbf{Z}_{t-1} + \mathbf{\Phi}\mathbf{s}_t,$$

and the observation equation is

$$\mathbf{x}_t = \mathbf{H}\mathbf{Z}_t,$$

In addition, the eigenvalues of $\mathbf{F}$ *all lie inside the unit circle.*

Proof. See Appendix A.2 for proof. □

This lemma states that any finite ARMA process has at least one state-space representation. However, it is often the case that there are many different state-state representations for the same ARMA process. More generally, the state equation can be written as

$$\mathbf{Z}_t = \mathbf{F}\mathbf{Z}_{t-1} + \mathbf{\Phi}\mathbf{s}_t,$$

and the observation equation can be written as

$$\mathbf{x}_t = \mathbf{H}\mathbf{Z}_t + \mathbf{\Psi}\mathbf{s}_t.$$

Lemma 3.1 only provides one of the state-space representation with the feature that there is no shock in the observation equation. Even though the state space representation is not unique, all different representations lead to the same forecast eventually.

Suppose that there exits $\mathbf{B}(L)$ and $\{\mathbf{w}_t\}$ such that

$$\mathbf{x}_t = \mathbf{M}(L)\mathbf{s}_t = \mathbf{B}(L)\mathbf{w}_t, \tag{3.9}$$

where $\mathbf{B}(L)$ is invertible,[7] and $\mathbf{w}_t$ is serially uncorrelated shocks with co-variance matrix $\mathbf{V}$, then we say $\mathbf{x}_t = \mathbf{B}(L)\mathbf{w}_t$ is a fundamental representation of $\mathbf{x}_t$. The Wold representation is a particular fundamental representation. Since $\mathbf{B}(L)$ is invertible, $\mathbf{x}^t$ contains the same information as $\mathbf{w}^t$, i.e., $\mathcal{H}_t^x = \mathcal{H}_t^w$. Further, equation (3.9) implies that the auto-correlation generating function can be obtained using both representations

$$\boldsymbol{\rho}_{xx}(z) = \mathbf{M}(z)\mathbf{M}'(z^{-1}) = \mathbf{B}(z)\mathbf{V}\mathbf{B}'(z^{-1}). \tag{3.10}$$

Therefore, find the fundamental representation is equivalent to find the canonical factorization of the auto-correlation generating function . The following theorem provides the canonical factorization for the state-space representation of the signal process $\mathbf{x}_t$, which uses the factorization identity.

Theorem (Canonical Factorization)**.** *Let* $\mathbf{F}$ *denote an matrix whose eigenvalues are all inside the unit circle; let* $\mathbf{Q}'\mathbf{Q}$ *be positive definite matrix; let* $\mathbf{H}$ *denote an arbitrary matrix. Let* $\mathbf{P}$ *satisfy*

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

and $\mathbf{K}$ *be defined as*

$$\mathbf{K} = \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \boldsymbol{\Psi}\boldsymbol{\Psi}')^{-1}$$

Then

1. *The eigenvalues of* $\mathbf{F} - \mathbf{F}\mathbf{K}\mathbf{H}$ *are all inside the unit circle.*

2. *The canonical factorization is*

$$\boldsymbol{\rho}_{xx}(z) = \mathbf{H}[\mathbf{I} - \mathbf{F}z]^{-1}\boldsymbol{\Phi}\boldsymbol{\Phi}'[\mathbf{I} - \mathbf{F}z^{-1}]^{-1}\mathbf{H}'$$

[7] This is equivalent to that the determinant of $\mathbf{B}(z)$ does not contain any roots (zeros) within the unit circle.

$$= [\mathbf{I} + \mathbf{H}(\mathbf{I} - \mathbf{F}z)^{-1}\mathbf{FK}z][\mathbf{HPH}' + \mathbf{\Psi\Psi}'][\mathbf{I} + \mathbf{K}'\mathbf{F}'(\mathbf{I} - \mathbf{F}'z^{-1})^{-1}\mathbf{H}'z^{-1}]$$
$$\equiv \mathbf{B}(z)\mathbf{VB}'(z^{-1}).$$

3. *$\mathbf{B}(z)$ is*

$$\mathbf{B}(z) = \mathbf{I} + \mathbf{H}[\mathbf{I} - \mathbf{F}z]^{-1}\mathbf{FK}z,$$

the inverse of $\mathbf{B}(z)$ is

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

and the co-variance matrix $\mathbf{V}$ is

$$\mathbf{V} = \mathbf{HPH}' + \mathbf{\Psi\Psi}'$$

Proof. The proof is in Hamilton (1994). □

The proof utilizes the forecasting formula in the steady-state Kalman filter. The requirement that all the eigenvalues of $\mathbf{F}$ lie inside the unit circle guarantees $\mathbf{I}-\mathbf{F}z$ is invertible. The eigenvalues of $\mathbf{F} - \mathbf{FKH}$ are related to the Kalman gains and are very important in understanding the prediction problem, which essentially determines the persistence of the forecasts.

3.3 Wiener-Hopf prediction formula

Now we turn to the inference problems incorporated in equation (3.4). The following theorem states the Wiener-Hopf prediction formula. Note that this prediction formula does not hinge on whether the signal follows a finite ARMA process or not.

Theorem (Wiener-Hopf)**.** *Suppose the multivariate co-variance stationary signal process follows*

$$\mathbf{x}_t = \mathbf{M}(L)\mathbf{s}_t,$$

and f_t is a univariate co-variance stationary process

$$f_t = \boldsymbol{\psi}(L)\mathbf{s}_t.$$

where $\boldsymbol{\psi}(L) = \sum_{k=-\infty}^{\infty} \boldsymbol{\psi}_k L^k$ is a non-casual stationary lag polynomial. Then the optimal *linear prediction of f_t conditional on $\{\mathbf{x}_t\}$ is*

$$\mathbb{E}\left[f_t|\mathbf{x}^t\right] = \left[\boldsymbol{\psi}(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_+ \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{x}_t, \tag{3.11}$$

where $\mathbf{B}(L)$ and $\mathbf{V}$ are given by the Canonical Factorization Theorem.

Proof. See Appendix A.3 for proof. □

If we further assume that the signal follows a finite ARMA process, we can obtain a sharper and more specific prediction formula.

Proposition 3.1. *Under Assumption 3.1,*

$$\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} = \frac{1}{\Pi_{k=1}^{u}(L-\lambda_k)}\mathbf{G}(L) \tag{3.12}$$

where $\mathbf{G}(L)$ is a polynomial matrix in L, and $\{\lambda_k\}_{k=1}^{u}$ are non-zero eigenvalues of $\mathbf{F}-\mathbf{FKH}$ which all lie inside the unit circle. Suppose the univariate random variable f_t follows

$$f_t = \frac{\mathbf{a}(L)}{\prod_{\tau=1}^{d}(L-\beta_\tau)}\mathbf{s}_t,$$

where $\mathbf{a}(L)$ is a casual stationary polynomial. The prediction formula for f_t is [8]

$$\begin{aligned}
\mathbb{E}\left[f_t \mid x^t\right] \quad = \quad & \frac{\mathbf{a}(L)}{\prod_{\tau=1}^{d}(L-\beta_\tau)}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\mathbf{x}_t & (3.13)\\
& -\sum_{k=1}^{u}\frac{\mathbf{a}(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(L)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)}\mathbf{x}_t & (3.14)\\
& -\sum_{k=1}^{d}\frac{\mathbf{a}(\beta_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(L)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\mathbf{x}_t & (3.15)
\end{aligned}$$

Proof. See Appendix A.4 for proof. □

The key in applying the Wiener-Hopf prediction formula is to find the Wold representation for $\mathbf{x}_t$ or the canonical factorization for $\mathbf{M}(L)$. When the number of signals equals the number

[8] Here, we do not consider the case with repeated roots. It is straightforward to add those but make the expression cumbersome.

of shocks, $\mathbf{M}(L)$ is a square matrix. Suppose $\mathbf{M}(L)$ is invertible, then $\mathbf{M}(L)$ itself is a fundamental representation and the Wiener-Hopf prediction formula can be applied directly. This corresponds to the case when the signals fully reveal the underlying states. If $\mathbf{M}(L)$ is a square but not an invertible matrix, then there exists at least one inside root of the determinant of $\mathbf{M}(L)$. In this case, the fundamental representation can be found by multiplying the Blaschke matrices to flip the inside roots outside the unit circle. Kasa (2000), Rondina and Walker (2017), Kasa, Walker, and Whiteman (2014) and Acharya (2013) all use this method to find the Wold representation.

In most signal extraction problems, the number of shocks is larger than the number of signals. In this case, $\mathbf{M}(L)$ is a non-square matrix and is not invertible. To find the canonical factorization of $\mathbf{M}(L)$ is more involved, but can be achieved by using the Canonical Factorization Theorem.

It is common in the existing literature to restrict the number of signals to being the same as the number of shocks so that the Blaschke matrix is applicable in finding the Wold representation. This is mainly because a square system can be helpful to obtain analytic results when information is endogenous. However, this restriction often leaves some informative variables to be observed without noise. As a result, the true state of the economy is revealed too quickly. For example, Kasa (2000), Sargent (1991) and Pearlman and Sargent (2005) all show that in Townsend (1983), agents share the same belief about the common demand shock and there is no *forecast the forecasts of others* problem. Also, the forecast error only exists for one period, and agents figure out the demand shock fairly quickly. The one period delay is due to the fact that output is predetermined. Similarly, in Acharya (2013), agents observe the last period's aggregate output perfectly, and effects of aggregate noise only last for one period because agents can infer the underlying shock accurately by observing aggregate output. Rondina and Walker (2017) and Kasa, Walker, and Whiteman (2014) both have square observation matrix. To prevent the price from fully revealing the information, they need to abandon the standard AR(1) process and assume a non-invertible process for the fundamental.

A lot of interesting models naturally require that there are more shocks than signals, such as Singleton (1987), Woodford (2003), Lorenzoni (2009), Angeletos and La'O (2010), Angeletos and La'O (2013) and so on. In this paper, we show that by using the factorization identity, the Wold representation is readily available for any finite ARMA process. Joint with the Wiener filter, we can solve the signal extraction problem.

3.4 System of analytic functions and a finite-state representation

To write it more compactly for future derivation, define

$$\boldsymbol{\phi}(L) \equiv \mathbf{vec}(\mathbf{h}(L)). \tag{3.16}$$

$\boldsymbol{\phi}(L)$ effectively collapse all the lag polynomials to be solved into a $rn \times 1$ vector.

After we apply the Wiener filter, solving for $\boldsymbol{\phi}(L)$ in model (3.4) still requires solving sequences of infinite coefficients in the lag polynomials, which is an infinite dimension problem. By the Riesz-Fisher Theorem, instead of solving the sequences of infinite coefficients, we can solve for a finite number of analytic functions instead, as shown in the following proposition.

Proposition 3.2. *Under Assumption 3.1 and 3.2, there exists a solution $\boldsymbol{\phi}(L)$ to model (3.4) if and only if there exists a vector analytic function $\boldsymbol{\phi}(z)$ that solves*

$$\mathbf{T}(z)\boldsymbol{\phi}(z) = \mathbf{D}\bigg[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\bigg] \tag{3.17}$$

where $\mathbf{T}(z)_{rn\times rn}$ is given by

$$\mathbf{T}(z) \equiv \mathbf{P}(z) \otimes (\mathbf{M}(z^{-1})\mathbf{M}'(z)) + \mathbf{Q}(z) \otimes (\mathbf{M}(z^{-1})\boldsymbol{\Lambda}\mathbf{M}'(z)) \tag{3.18}$$

and $\mathbf{D}\bigg[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\bigg]_{rn\times 1}$ is given by

$$\begin{aligned}
&\mathbf{D}\bigg[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\bigg] \equiv -\mathbf{vec}(\mathbf{M}(z^{-1})\mathbf{R}'(z)) \\
&+\sum_{k=1}^{u} \frac{\mathbf{vec}(\mathbf{B}(z^{-1})\mathbf{G}'(\lambda_k)\widehat{\mathbf{R}}'(\lambda_k))}{(z-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{d} \frac{\mathbf{vec}(\mathbf{B}(z^{-1})\mathbf{G}'(\beta_k)\widehat{\mathbf{R}}'(\beta_k))}{(z-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\
&+(\mathbf{I}_r \otimes \mathbf{B}(z^{-1}))\sum_{k=1}^{u} \frac{\widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\boldsymbol{\Lambda}\mathbf{M}'(\lambda_k)}{(z-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)}\boldsymbol{\phi}(\lambda_k) \\
&+(\mathbf{I}_r \otimes \mathbf{B}(z^{-1}))\sum_{k=1}^{d} \frac{\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k) + \widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\boldsymbol{\Lambda}\mathbf{M}'(\beta_k)}{(z-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\boldsymbol{\phi}(\beta_k)
\end{aligned} \tag{3.19}$$

Proof. See Appendix A.5 for proof. □

To solve for $\boldsymbol{\phi}(z)$, we utilize the Cramer's rule. Different from solving a system of linear equations, one needs to make sure $\boldsymbol{\phi}(z)$ represents a stationary process. To proceed, we first determine the set of constants $\{\{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\}$ that are generated when applying the Wiener-Hopf prediction formula. As discussed in Whiteman (1983), these constants should be set to remove the poles of $\boldsymbol{\phi}(z)$ that are inside the unit circle, which makes sure that $\boldsymbol{\phi}(z)$ is an analytic function. The number of free constants that can be used to eliminate the poles of $\boldsymbol{\phi}(z)$ may be smaller than the cardinality of $\{\{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\}$, because some of these constants enter equation (3.17) in a linearly dependent way. The following lemma gives the actual number of free constants.

Lemma 3.2. *Define N_1 as*

$$N_1 = \sum_{k=1}^{u} \mathbf{rank}\Big(\widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\boldsymbol{\Lambda}\mathbf{M}'(\lambda_k)\Big) + \sum_{k=1}^{d} \mathbf{rank}\Big(\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k) + \widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\boldsymbol{\Lambda}\mathbf{M}'(\beta_k)\Big).$$

There exists a $N_1 \times 1$ constant vector $\boldsymbol{\psi}$ such that

$$\mathbf{D}\Big[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\Big] \quad = \quad \mathbf{D}_1(z)\boldsymbol{\psi} + \mathbf{D}_2(z), \tag{3.20}$$

where $\mathbf{D}_1(z)$ is with full column rank and $\boldsymbol{\psi}$ is a linear combination of $(\{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d})$.

Proof. See Appendix A.6 for proof and the construction of $\mathbf{D}_1(z)$, $\mathbf{D}_2(z)$, and $\boldsymbol{\psi}$. □

The roots of $\mathbf{T}(z)$ are crucial in obtaining the solution and in understanding its property.

1. Denote $\{\zeta_1, \ldots, \zeta_{N_2}\}$ as the N_2 roots of $\det[\mathbf{T}(z)]$ that lie inside the unit circle.
2. Denote $\{\vartheta_1^{-1}, \ldots, \vartheta_{N_3}^{-1}\}$ as the N_3 roots of $\det[\mathbf{T}(z)]$ that lie outside the unit circle.

The existence of a solution hinges on whether there are enough free constants to eliminate all the inside roots of $\det[\mathbf{T}(z)]$. The persistence of $\boldsymbol{\phi}(z)$ depends on the outside roots of $\det[\mathbf{T}(z)]$.

Theorem 1 (General solution formula)**.** *Assume Assumption 3.1 and 3.2 hold.* $\mathbf{U}_1$ *and* $\mathbf{U}_2$

are constructed such that

$$\mathbf{U}_1\boldsymbol{\psi} + \mathbf{U}_2 \equiv \begin{bmatrix} \det\begin{bmatrix}\mathbf{D}_1(\zeta_1)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_1) & \mathbf{T}_2(\zeta_1) & \dots & \mathbf{T}_{rn}(\zeta_1)\end{bmatrix} \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ \det\begin{bmatrix}\mathbf{D}_1(\zeta_{N_2})\boldsymbol{\psi} + \mathbf{D}_2(\zeta_{N_2}) & \mathbf{T}_2(\zeta_{N_2}) & \dots & \mathbf{T}_{rn}(\vartheta_{N_2})\end{bmatrix} \end{bmatrix}.$$

1. *If $N_1 < N_2$, there is no solution.*
2. *If $N_1 = N_2 = \mathbf{rank}(\mathbf{U}_1)$, there exists a unique solution $\boldsymbol{\phi}(z)$. For $i \in \{1, \dots, rn\}$*

$$\phi_i(z) = \frac{\det\begin{bmatrix}\mathbf{T}_1(z) & \dots & \mathbf{T}_{i-1}(z) & \mathbf{D}_1(z)\boldsymbol{\psi} + \mathbf{D}_2(z) & \mathbf{T}_{i+1}(z)\dots & \dots & \mathbf{T}_{rn}(z)\end{bmatrix}}{\det\begin{bmatrix}\mathbf{T}(z)\end{bmatrix}},$$

and

$$\boldsymbol{\psi} = -\mathbf{U}_1^{-1}\mathbf{U}_2,$$

3. *If $N_1 > N_2$ or $N_1 = N_2 > \mathbf{rank}(\mathbf{U}_1)$, there exists an infinite number of solutions.*

Proof. See Appendix A.7 for proof. □

Theorem 1 gives the existence and uniqueness condition for the solution. It also delivers the exact formula of the solution. The following proposition further shows that the solution can always be represented in a recursive way.

Proposition 3.3 (Finite-state representation)**.** *Under Assumption 3.1 and 3.2, if there exists a solution $\boldsymbol{\phi}(L)$ to model (3.4), then $\boldsymbol{\phi}(L)$ is a rational function of lag operator L*

$$\boldsymbol{\phi}(L) = \frac{\widehat{\boldsymbol{\phi}}(L)}{\prod_{k=1}^{N_3}(1 - \vartheta_k L)} \tag{3.21}$$

where $\widehat{\boldsymbol{\phi}}(L)$ is a lag polynomial matrix with finite degree.

Given a particular signal realization $\{\mathbf{x}_t\}_{t=-\infty}^{-1}$, there exists a finite set of state variables $\mathbf{z}_t$, *such that the choice variables $\mathbf{y}_t$ have a finite-state representation*

$$\mathbf{y}_t = \mathbf{\Gamma}_z \mathbf{z}_{t-1} + \mathbf{\Gamma}_x \mathbf{x}_t, \tag{3.22}$$

where the law of motion of $\mathbf{z}_t$ and the initial state are given by

$$\mathbf{z}_t = \mathbf{\Upsilon}_z \mathbf{z}_{t-1} + \mathbf{\Upsilon}_x \mathbf{x}_t, \tag{3.23}$$

$$\mathbf{z}_{-1} = (\mathbf{I} - \mathbf{\Upsilon}_z L)^{-1} \mathbf{\Upsilon}_x \mathbf{x}_{-1}. \tag{3.24}$$

Proof. See Appendix A.8 for proof. □

This result implies that when signals follow finite ARMA processes, higher order beliefs do not create infinite state variables. It is always possible to use a small set of variables to summarize the necessary information in the past. An important property is that the persistence of the policy rule is completely characterizes by the roots of the determinant of $\mathbf{T}(z)$, and these roots are determined by the interaction between the model primitives ($\mathbf{P}(L)$ and $\mathbf{Q}(L)$) and the signal process ($\mathbf{M}(L)$). This is in contrast with common information models.

Comparison with Common Information Case The special cases discussed in subsection 3.1 have the feature that all agents' signals are driven by the same shocks, and it follows that $\mathbf{\Lambda} = \mathbf{I}$. The matrix $\mathbf{T}(z)$ in equation (3.18) turns out to be

$$\mathbf{T}(z) = (\mathbf{P}(z) + \mathbf{Q}(z)) \otimes (\mathbf{M}(z^{-1})\mathbf{M}'(z)), \tag{3.25}$$

and its determinant is

$$\det[\mathbf{T}(z)] = \det[\mathbf{P}(z) + \mathbf{Q}(z)] \ \det[\mathbf{M}(z^{-1})\mathbf{M}'(z)] \tag{3.26}$$

The roots $\{\vartheta_1^{-1}, \ldots, \vartheta_{N_3}^{-1}\}$ that determines the persistence of the policy rule either come from the primitive part of the model $\mathbf{P}(z)+\mathbf{Q}(z)$ or from the signal process $\mathbf{M}(z^{-1})\mathbf{M}'(z)$. Crucially, there is no interaction between the model and the signal in determining $\{\vartheta_k\}$. In contrast, when $\mathbf{\Lambda} \neq \mathbf{I}$, the determinant of $\mathbf{T}(z)$ is jointly determined by $\mathbf{P}(z), \mathbf{Q}(z)$, and $\mathbf{M}(z)$, which is what separates the dispersed information model from the representative agent model.

Innovation form and signal form The solution we discussed in Section 3.4 is in terms of signals, and it can also be represented in terms of the underlying shocks $\{\mathbf{s}_t\}$

$$\mathbf{y}_t = \mathbf{d}(L)\mathbf{s}_t. \tag{3.27}$$

We label the solution in terms of signals as *signal form* and the solution in terms of underlying shocks as *innovation form*. On one hand, the signal form is typically easier to solve, because the dimension of the problem in signal form is smaller than the dimension of the problem in innovation form. On the other hand, it is more straightforward to characterize the equilibrium using the innovation form. The following proposition shows that one can work with either of them.

Proposition 3.4. *Under Assumption 3.1 and 3.2, there exists a solution in signal form,*

$$\mathbf{y}_t = \mathbf{h}(L)\mathbf{x}_t, \tag{3.28}$$

if and only if there exists a solution in innovation form

$$\mathbf{y}_t = \mathbf{d}(L)\mathbf{s}_t, \tag{3.29}$$

where $\mathbf{h}(L)$ *and* $\mathbf{d}(L)$ *satisfy*

$$\mathbf{d}(L) = \mathbf{h}(L)\mathbf{M}(L),$$
$$\mathbf{vec}(\mathbf{h}(L)) = \mathbf{vec}(\boldsymbol{\rho}'_{xx}(L)^{-1}\mathbf{M}(L^{-1})\mathbf{d}'(L)) - \sum_{k=1}^{u} \frac{\mathbf{vec}(\mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\lambda_k)\mathbf{d}'(\lambda_k))}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)}.$$

Proof. See Appendix A.9 for proof. $\square$

If $\mathbf{M}(L)$ is not invertible, the space spanned by signals is a subset of the space spanned by shocks. It should be clear that whether we use the innovation form or the signal form, $\{\mathbf{y}_t\}$ always lies in the space spanned by current and past signals because agents can only condition their choice on their observables, that is, $\{\mathbf{y}_t\} \subset \mathcal{H}_t^x \subset \mathcal{H}_t^s$.

4 Application

In this section, we use the method developed in Section 3 to solve some stylized problems. These problems are commonly adopted in the literature, and are useful in illustrating how to use our method.

4.1 Application I: Static Beauty Contest Model

We first consider the static beauty contest model. A continuum of agents choose their actions according to the following best response function

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_t], \tag{4.1}$$

and the aggregate action y_t is given by

$$y_t = \int y_{it}. \tag{4.2}$$

This type of best response can be derived directly from a quadratic utility function as in Morris and Shin (2002), and can be the result of linearized macro models as in Woodford (2003) or Angeletos and La'O (2010). We label this type of best response as a static beauty contest model for the reason that past or future actions do not enter agents' best response function. In equation (3.4), this model structure implies that $\mathbf{P}(L)$ and $\mathbf{Q}(L)$ are constant matrices instead of lag polynomial matrices.

Now we introduce the shocks and signals. Suppose that the economic fundamental ξ_t follows an AR(1) process

$$\xi_t = \rho \xi_{t-1} + \sigma_\eta \eta_t. \tag{4.3}$$

Agents do not observe the fundamental perfectly. They receive two signals about the fundamental. The first signal is a public signal observed by all agents, and the second signal is private that can only be observed by agent i.

$$x_{it}^1 = \xi_t + \sigma_\epsilon \epsilon_t, \quad x_{it}^2 = \xi_t + \sigma_u u_{it}.$$

More compactly, the signal structure can be represented as

$$\mathbf{x}_{it} = \mathbf{M}(L)\mathbf{s}_{it} = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & \sigma_\epsilon & 0 \\ \frac{\sigma_\eta}{1-\rho L} & 0 & \sigma_u \end{bmatrix} \begin{bmatrix} \eta_t \\ \epsilon_{it} \\ u_{it} \end{bmatrix}, \quad \mathbf{s}_{it} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{4.4}$$

For the inference problem, the corresponding state-space representation is

$$z_{it} = \xi_t = \underbrace{\rho}_{\mathbf{F}} \xi_{t-1} + \underbrace{\begin{bmatrix} \sigma_\eta & 0 & 0 \end{bmatrix}}_{\mathbf{\Phi}} \begin{bmatrix} \eta_t \\ \epsilon_{it} \\ u_{it} \end{bmatrix}, \qquad \mathbf{x}_{it} = \underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{\mathbf{H}} z_{it} + \underbrace{\begin{bmatrix} 0 & \sigma_\epsilon & 0 \\ 0 & 0 & \sigma_u \end{bmatrix}}_{\mathbf{\Psi}} \begin{bmatrix} \eta_t \\ \epsilon_{it} \\ u_{it} \end{bmatrix}. \tag{4.5}$$

Besides the best response function, this information structure also resembles those in Morris and Shin (2002), Woodford (2003), and Angeletos and La'O (2010). Morris and Shin (2002) study how the precision of the public signal σ_ϵ affects the social welfare, while Angeletos and La'O (2010) interpret ϵ_t as animal spirits and study how these shocks shape aggregate fluctuations. In Woodford (2003), agents only receive a private signal about the fundamental, which is equivalent to set σ_ϵ to infinity. Woodford (2003) and Angeletos and La'O (2010) use a guess-and-verify approach to solve the model, and we provide the analytic solution in a systematic way.

Given the equilibrium policy rule $y_{it} = \mathbf{h}(L)\mathbf{x}_{it}$, the aggregate action is

$$y_t = \int \mathbf{h}(L)\mathbf{M}(L)\mathbf{s}_{it} = \mathbf{h}(L)\mathbf{M}(L)\mathbf{\Lambda}\mathbf{s}_{it}, \tag{4.6}$$

where the matrix that selects the common shocks η_t and ϵ_t is given by

$$\mathbf{\Lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{4.7}$$

In the static beauty contest model, $P(L) = 1$, $Q(L) = -\alpha$ and $\mathbf{R}(L) = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & 0 & 0 \end{bmatrix}$. By Proposition 3.3, the inverse of the persistence of the policy rule ϑ, is the outside root of $\det[\mathbf{T}(z)]$, where $\mathbf{T}(z)$ is given by

$$\mathbf{T}(z) = \mathbf{M}(z)(\mathbf{I} - \alpha\mathbf{\Lambda})\mathbf{M}'(z^{-1}). \tag{4.8}$$

So far, all the elements to be used in Proposition 3.2 are at hand. The following proposition specifies the finite-state representation of the equilibrium policy rule

Proposition 4.1. *Given the signal process (4.4) and $\alpha \in (0,1)$, the equilibrium policy rule in model (4.1) is given by*

$$h_1(L) = \frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \tag{4.9}$$

$$h_2(L) = \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \tag{4.10}$$

where

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

The finite-state representation is

$$y_{it} = \vartheta y_{it-1} + \frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}x_{it}^1 + \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)}x_{it}^2 \tag{4.12}$$

$$y_t = \vartheta y_{t-1} + \frac{\vartheta}{\rho(1-\rho\vartheta)}\frac{(1-\alpha)\sigma_\epsilon^2+\sigma_u^2}{(1-\alpha)\sigma_\epsilon^2\sigma_u^2}\xi_t + \frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}\sigma_\epsilon\epsilon_t \tag{4.13}$$

Proof. See Appendix A.10 for proof. □

The individual policy rule follows an AR(1) process, which inherits the property of the underlying fundamental. As expected, the weights assigned to the signals are adjusted according to their relative informativeness. In addition, the first signal affects all agents in the economy, each individual agent will respond to it more strongly. As the strength of the strategic complementarity increases (α increases), the instantaneous response to the first signal, $\frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}$, also becomes larger.

Crucially, the persistence of agents' endogenous action is governed by the endogenous variable ϑ, and it is the key to understand the nature of the equilibrium. Woodford (2003) emphasizes that higher order beliefs generate inertia of agents' action, which displays a hump-shaped response to the fundamental shock. In equation (4.11), given ρ, as ϑ increases, the peak of the impulse response of y_t shifts to the right. If ϑ is small enough, then hump-shaped response may disappear. The following proposition characterizes ϑ.

Proposition 4.2. *Assume that $\alpha \in (0,1), \rho \in (0,1), \sigma_\epsilon > 0$, and $\sigma_u > 0$. Then ϑ satisfies*

1. $0 < \lambda < \vartheta < \rho$, *where λ is given by*

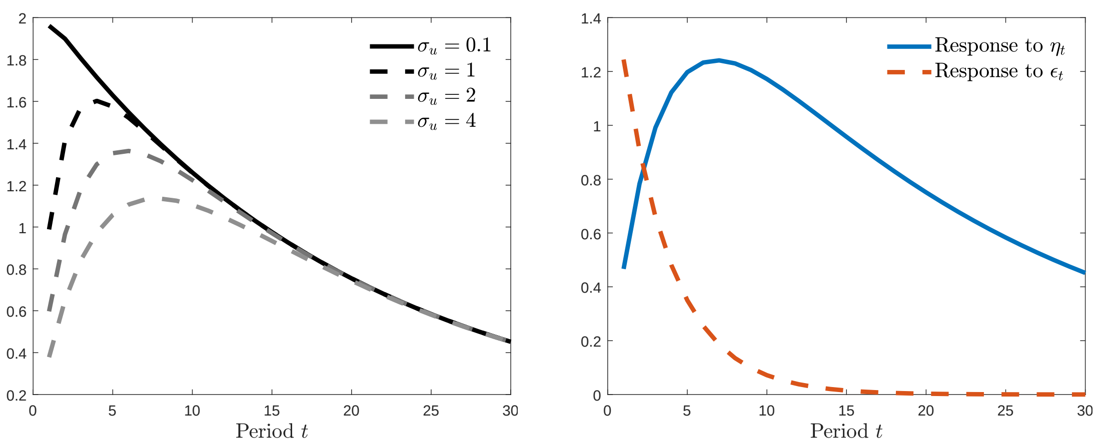
$$\lambda = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{\sigma_\epsilon^2 + \sigma_u^2}{\rho\sigma_\epsilon^2\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\sigma_\epsilon^2 + \sigma_u^2}{\rho\sigma_\epsilon^2\sigma_u^2}\right)^2 - 4}\right] \tag{4.14}$$

2. ϑ *is increasing in α and*

$$\lim_{\alpha \to 0} \vartheta = \lambda$$

3. ϑ *is increasing in σ_ϵ, σ_u, and ρ.*

Here, ϑ is bounded from above by the persistence of ξ_t, and it is also bounded from below by λ, where $1 - \lambda$ is the Kalman gain when using the Kalman filter to predict ξ_t. Note that ϑ is increasing in α. This is because higher order beliefs become more important in shaping the behavior of y_t. With a large α, it is more likely for y_t to have a hump-shaped response.



(A) Response to η_t shock (B) Response to η_t and ϵ_t shock

FIGURE 1: Impulse Response of Aggregate Action

Parameters: $\alpha = 0.5, \rho = 0.95, \sigma_\eta = 1, \sigma_\epsilon = 4$.

We use a numerical example to illustrate the properties of the model when varying the degree information frictions. We choose different values for the variance of the idiosyncratic shock σ_u^2. As shown in Figure 1a, the hump-shaped response of y_t to η_t shock is more pronounced when information frictions becomes larger. Figure 1b compare the responses of y_t to the fundamental shock and the common noise shock. The response to the common noise shock follows an AR(1) process, and the persistence is determined by ϑ. These patterns are in line with the findings in Woodford (2003) and Angeletos and La'O (2010).

4.2 Application II: Dynamic Beauty Contest Model

In this section, we explore the beauty contest modes with dynamic consideration in their payoff structure. For example, when consumers making a saving-consumption decision, or when firms setting their prices subject to nominal rigidities, they need to look forwards or backwards to solve their optimization problems. Particularly, we compare three types of models.

1. Forward-looking model, with $Q(L) = -\alpha L^{-1}$

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_{t+1}]. \tag{4.15}$$

2. Static model, with $Q(L) = -\alpha$

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_t]. \tag{4.16}$$

3. Backward-looking model, with $Q(L) = -\alpha L$

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_{t-1}]. \tag{4.17}$$

As in subsection 4.1, we still assume that the fundamental follow an AR(1) process (4.3). In order to highlight how the persistence of the policy rule varies in these models, we assume that agents only observe one private signal about the fundamental

$$x_{it} = \xi_t + \sigma_u u_{it}. \tag{4.18}$$

The following proposition gives the analytical solution to these models

Proposition 4.3. *Given the signal process (4.18) and assume that $\alpha \in (0, 1)$, the equilibrium policy rules are*

1. *Forward-looking model*

$$y_{it} = \vartheta_f y_{it-1} + \frac{\sigma_\eta^2}{\sigma_u^2} \frac{\vartheta_f}{\rho\left(1 - \rho\vartheta_f - \alpha\frac{\sigma_\eta^2}{\sigma_u^2}\vartheta_f\right)} x_{it} \tag{4.19}$$

$$\textit{with} \quad \vartheta_f = \frac{1}{2\left(1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right)} \left[\left(\frac{1}{\rho} + \rho + \frac{\sigma_\eta^2}{\rho\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\sigma_\eta^2}{\rho\sigma_u^2}\right)^2 - 4\left(1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right)}\right]$$

2. *Static model*

$$y_{it} = \vartheta_s y_{it-1} + \frac{\sigma_\eta^2}{\sigma_u^2} \frac{\vartheta_s}{\rho(1 - \rho\vartheta_s)} x_{it} \tag{4.20}$$

$$\textit{with} \quad \vartheta_s = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\eta^2}{\rho\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\eta^2}{\rho\sigma_u^2}\right)^2 - 4}\right]$$

3. *Backward-looking model*

$$y_{it} = \vartheta_b y_{it-1} + \frac{\vartheta_b\sigma_\eta^2}{\rho\sigma_u^2 + \alpha\sigma_\eta^2 - \rho^2\sigma_u^2\vartheta_b} x_{it} \tag{4.21}$$

$$\textit{with} \quad \vartheta_b = \frac{1 + \frac{\alpha\sigma_\eta^2}{\rho\sigma_u^2}}{2}\left[\left(\frac{\sigma_\eta^2 + \sigma_u^2(1+\rho^2)}{\rho\sigma_u^2 + \alpha\sigma_\eta^2}\right) - \sqrt{\left(\frac{\sigma_\eta^2 + \sigma_u^2(1+\rho^2)}{\rho\sigma_u^2 + \alpha\sigma_\eta^2}\right)^2 - 4\left(\frac{\rho\sigma_u^2}{\rho\sigma_u^2 + \alpha\sigma_\eta^2}\right)}\right]$$

Furthermore, the ranking of the persistence is

$$\vartheta_b > \vartheta_s > \vartheta_f. \tag{4.22}$$

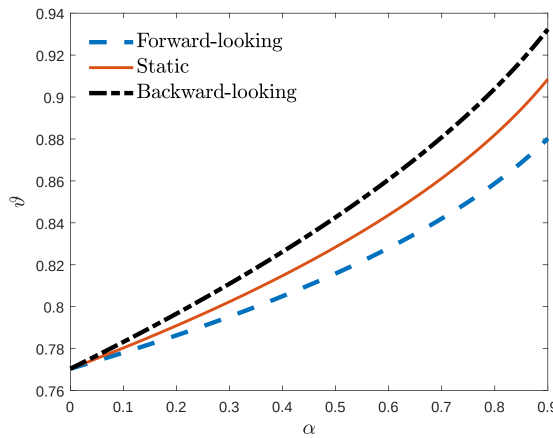
Proof. See Appendix A.11. □

This proposition first confirms that the finite-state property of the equilibrium policy rule still holds with dynamic beauty contest models, and the complexity of the required state variables is similar to the static model. In terms of the persistence of the policy rule, they all depend on the degree of complementarity α, but different models are affected by α in different ways. The backward-looking model exhibits the highest persistence and the forward-looking model exhibits the lowest persistence. Intuitively, when agents care more about the past aggregate action, they will naturally assign a higher weight on past signals.

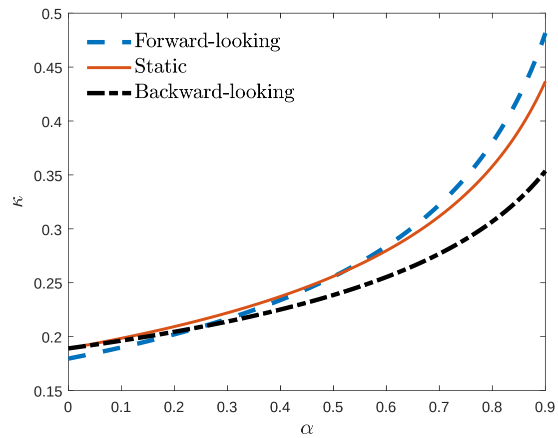
The aggregate action in all the three models can be represented as

$$y_t = \vartheta y_{t-1} + \kappa \xi_t.$$

Figure 2 shows that how the persistence ϑ and the instantaneous response κ change as the degree of complementarity α changes. In all three models, the persistence ϑ is increasing in α, and the backward-looking model always has the highest persistence. The instantaneous response also increases with α, but the ranking across the three models changes with different values of α.



(A) Persistence



(B) Instantaneous Response

FIGURE 2: Effects of α on Aggregate Action

Parameters: $\rho = 0.95, \sigma_\eta = 1, \sigma_u = 4$.

4.3 Application III: Independent Value Model with Random Matching

Applicaiton I and II consider the case where each agent interacts with the average of a continuum of other agents. In this section, we adopt the model environment in Angeletos and La'O (2013) in which an agent meets a different player every period. In the baseline model of Angeletos and La'O (2013), it is assumed that there is no persistent shock. This assumption does not affect their qualitative prediction, and it helps to avoid the infinite regress problem. However, this assumption prevents the model from exploring more relevant learning problems. Here, we extend Angeletos and La'O (2013) to allow persistent shocks in the model.

Assume that there is a continuum of agents in the economy. An individual agent i is endowed with a permanent fundamental a_i, drawn from a normal distribution $\mathcal{N}(0, \sigma_a^2)$. 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

$$y_{it} = a_i + \alpha\, \mathbb{E}_{it}[y_{m(i,t)t}], \tag{4.23}$$

where α controls the degree of strategic complementarity, and $y_{m(i,t)t}$ is the action of agent i's match in period t. Angeletos and La'O (2013) provides the micro-foundation for this best response in a decentralized trading and production setting.

Besides her own fundamental, agent i receives two signals every period

$$x_{it}^1 = a_{m(i,t)} + \sigma_\epsilon \epsilon_{it}, \tag{4.24}$$

$$x_{it}^2 = x_{m(i,t)t}^1 + \xi_t + \sigma_u u_{it}, \tag{4.25}$$

where ϵ_{it} and u_{it} are both idiosyncratic noises, and ξ_t is a common noise. The fundamental of i's match is $a_{m(i,t)}$, and from i's perspective, it is also an i.i.d shock that follows $\mathcal{N}(0, \sigma_a^2)$. As emphasized by Angeletos and La'O (2013), agent i's forecast about $a_{m(i,t)}$ 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 beliefs are affected by the common noise ξ_t. In aggregate, ξ_t can generate aggregate flucutations, and can be interopreted as sentiments. In Angeletos and La'O (2013), ξ_t is an i.i.d shock, but we instead assume that ξ_t follows a persistent process.

$$\xi_t = \rho \xi_{t-1} + \sigma_\eta \eta_t. \tag{4.26}$$

In equilibrium, agent i's action is given by

$$y_{it} = h_a a_i + h_1(L) x_{it}^1 + h_2(L) x_{it}^2. \tag{4.27}$$

For this particular application, we need to solve for a constant h_a and two lag polynomials $h_1(L)$ and $h_2(L)$. Note that a_i is included for two reasons: first, it enters agent i's best response directly; second, a_i enters agent $m(i,t)$'s signal, and it also helps to predict i's match's action $y_{m(i,t)t}$.

Different from the applications discussed in subsection 4.1 and 4.2, agent i has to form higher order beliefs about a random player $m(i,t)$ every period. Our method developed in section 3 can still be applied to solve this model.

Proposition 4.4. *Assume that $\alpha \in (0,1)$. The finite-state representation of the equilibrium policy rule in model (4.23) is given by*

$$y_{it} = \vartheta y_{it-1} + h_a(1-\vartheta)a_i + \varphi(x^1_{it} - \vartheta x^1_{it-1}) + \frac{\alpha\varphi\vartheta}{\rho}\frac{\sigma^2_\epsilon\sigma^2_\eta}{\sigma^2_\epsilon + \sigma^2_u}(x^2_{it} - \rho x^2_{it-1}) \tag{4.28}$$

where

$$\vartheta = \frac{1}{2}\left[\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma^2_\eta}{\rho(\sigma^2_\epsilon + \sigma^2_u)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma^2_\eta}{\rho(\sigma^2_\epsilon + \sigma^2_u)}\right)^2 - 4}\right], \tag{4.29}$$

$$\varphi = \alpha\left(1 - \alpha^2 + \frac{\sigma^2_\epsilon}{\sigma^2_a}\left(1 - \alpha^2\frac{\vartheta}{\rho}\frac{\sigma^2_\epsilon\sigma^2_\eta}{\sigma^2_\epsilon + \sigma^2_u}\right)\right)^{-1}, \tag{4.30}$$

$$h_a = 1 + \alpha\varphi - \frac{\alpha\vartheta\varphi}{\rho}\frac{\sigma^2_\epsilon\sigma^2_\eta}{\sigma^2_\epsilon + \sigma^2_u}\frac{1-\rho}{1-\vartheta} \tag{4.31}$$

The aggregate y_t is given by

$$y_t = (\vartheta + \rho)y_{t-1} - \rho\vartheta y_{t-2} + \frac{\alpha\vartheta\varphi}{\rho}\frac{\sigma^2_\epsilon\sigma^2_\eta}{\sigma^2_\epsilon + \sigma^2_u}\eta_t \tag{4.32}$$

Proof. See Appendix A.13 for proof. □

Comparing with heterogeneous prior In the literature, a convenient device to avoid the infinite regress problem is to assume that agents have heterogeneous prior. 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 bias ξ_t. Given that agent i's policy rule is

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

then agent i believes that the action of her match is

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

In the heterogeneous prior equilibrium, the aggregate action is

$$y_t = \frac{\alpha^2}{(1-\alpha^2)(1-\alpha)}\xi_t. \tag{4.33}$$

Quantitatively, by assuming heterogeneous prior, y_t is perfectly correlated with ξ_t, while in our model with common prior, the persistence of aggregate action is endogenously determined by the structural parameter α and the information related parameters, and it is always different from the the persistence of ξ_t. A numerical example is shown in Figure 3. The solution

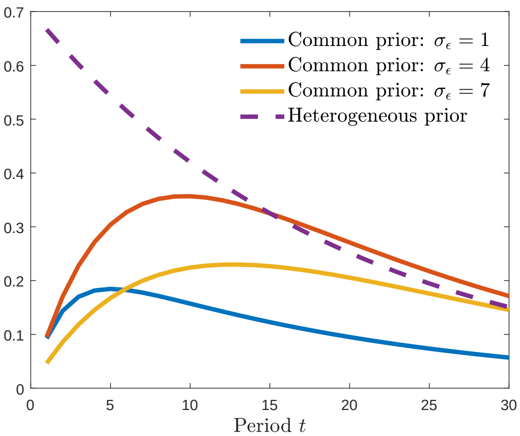


FIGURE 3: Effects of σ_ϵ on Aggregate Action

Parameters: $\rho = 0.95, \sigma_\eta = 1, \sigma_a = 3,\ \sigma_u = 1,\ \alpha = 0.5.$

under heterogeneous prior assumption is independent of the degree of information frictions by construction. The method we provide to solve the infinite regress problem retains the notion of rationality. Interestingly, by varying σ_ϵ, the aggregate action changes in a non-monotonic way. The size of the response maximizes when σ_ϵ takes an intermediate value.

4.4 Application IV: Heterogeneous Payoff Structure

Previous applications share a common feature which is all agents adopt the same best response and only differ from each other in terms of information sets. More generally, agents who interact with each other may be asymmetric in their payoff functions, and may depend on others' action to a different extend. A notable example is an economy with a network structure, where agents may care more about other agents who are closer to themselves. There has been a growing interest in these types of applications, and we show that our method is flexible enough to deal with them in this subsection.

Consider an economy with two agents. These two agents care about each other's action, but the degree of complimentary may not be the same. The best responses for agent 1 and 2 are

$$y_{1t} = (1-\omega_1)\mathbb{E}_{1t}[\xi_t] + \omega_1 \mathbb{E}_{1t}[y_{2t}], \tag{4.34}$$

$$y_{2t} = (1-\omega_2)\mathbb{E}_{2t}[\xi_t] + \omega_2 \mathbb{E}_{2t}[y_{1t}]. \tag{4.35}$$

This structure can be extended to a N-player economy in a straightforward way. In previous applications, a single parameter α is sufficient to summarize the dependence on others' action, while in the current setting with asymmetric players, it is necessary to use the matrix $\boldsymbol{\omega}$ to summarize the cross-dependence among different players

$$\boldsymbol{\omega} = \begin{bmatrix} 0 & \omega_1 \\ \omega_2 & 0 \end{bmatrix}. \tag{4.36}$$

If we interpret the the model economy as a network, this matrix $\boldsymbol{\omega}$ can also be thought as a weighted adjacency matrix.

For the information structure, we assume that each player i observes a noisy signal about the fundamental

$$x_{it} = \xi_t + \sigma_u u_{it}. \tag{4.37}$$

and the fundamental follows an AR(1) process (4.3). In equilibrium, agent i's action is $y_{it} = h_i(L)x_{it}$. Due to the fact that the two agents' best responses differ from each other, their equilibrium policy rules $h_1(L)$ and $h_2(L)$ will not be the same. For the special case in which $\omega_1 = \omega_2$, the two agents become symmetric and the solution is the same as the static beauty contest game in subsection 4.2.

The interaction between the adjacency matrix and the information frictions is our focus in this model economy. The following proposition characterizes this interaction.

Proposition 4.5. *Given the signal process (4.37) and assume that $|\omega_1\omega_2| \in (0,1)$, the equilibrium policy rules are*

$$h_1(L) = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_1\vartheta_2}{\rho(1-\vartheta_1\vartheta_2)}\frac{\frac{1}{\rho}\left((1-\omega_1)+\frac{(1-\omega_1\omega_2)}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)-\left((1-\omega_1)+\frac{(1-\omega_1\omega_2)\vartheta_1\vartheta_2}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)L}{(1-\vartheta_1 L)(1-\vartheta_2 L)}$$

$$h_2(L) = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_1\vartheta_2}{\rho(1-\vartheta_1\vartheta_2)}\frac{\frac{1}{\rho}\left((1-\omega_2)+\frac{(1-\omega_1\omega_2)}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)-\left((1-\omega_2)+\frac{(1-\omega_1\omega_2)\vartheta_1\vartheta_2}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)L}{(1-\vartheta_1 L)(1-\vartheta_2 L)}$$

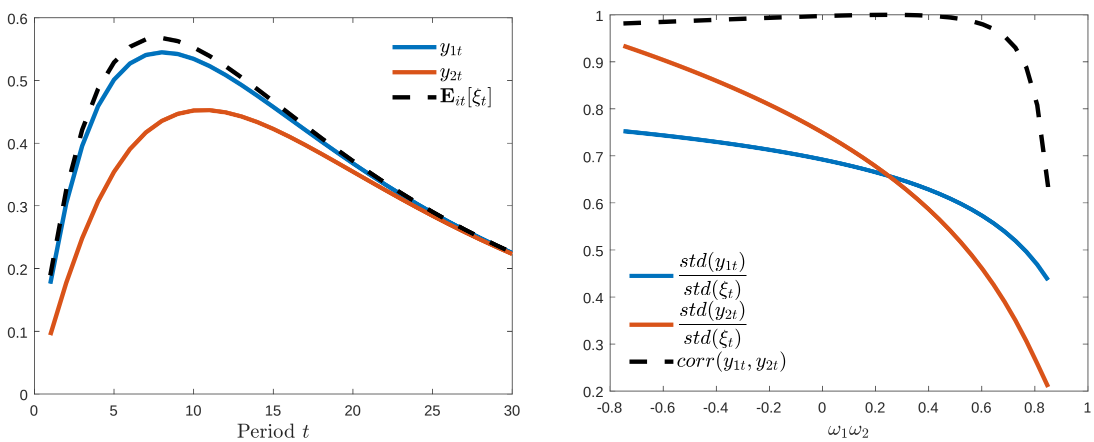
where

$$\vartheta_1 = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{(1-\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)^2 - 4}\right] \tag{4.38}$$

$$\vartheta_2 = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{(1+\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1+\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)^2 - 4}\right] \tag{4.39}$$

Proof. See Appendix A.12 for the proof. □

This proposition shows that the policy rules of both agents take an ARMA (2,1) form. Notably, the parameters ϑ_1 and ϑ_2 that determine the persistence of the actions are the same, which indicates that the actions of the two agents synchronize in the long run. Meanwhile, ω_1 and ω_2 do not enter ϑ_1 and ϑ_2 separately, and they affect ϑ_1 and ϑ_2 only through the product $\omega_1\omega_2$ which is propositional to the eigenvalues of $\boldsymbol{\omega}$. If $\omega_1 \neq \omega_2$, the responses of the actions in the short run will be different, captured by the MA part in the policy rules.



(A) IR to fundamental shock

(B) Impact of $\boldsymbol{\omega}$

Figure 4: Effects of τ_u on Aggregate Action

Parameters: $\rho = 0.95, \sigma_\eta = 1, \sigma_u = 4$. $\omega_1 = 0.1, \omega_2 = 0.9$ in panel A, $\omega_1 = 0.5$ in panel B

Figure 4 provides a numerical example. With $\omega_1 = 0.1$ and $\omega_2 = 0.9$, agent 1 wants to be close to the fundamental ξ_t, while agent 2 wants to be close to agent 1's action. As shown in panel A, after a shock to the fundamental, both agents have the same forecasts about the fundamental. Agent 1's action y_{1t} is close to the forecast about the fundamental, while agent

2's action is much more dampened due to the uncertainty about agent 1's action, and agent 1's belief about her own action, and so on. The difference in best response functions translates into the difference in actions, but the two agents' action synchronizes as time goes on. Panel B in Figure 4 shows how the actions vary with the degree of complementarities. We fix ω_1 to 0.5 and set ω_2 to different values. First, the volatilities of both agents' action decrease with $\omega_1\omega_2$. This is simply due to that higher order beliefs play a more important role when the degree of complementarity is large. Second, the correlation between the two agents' action peaks when $\omega_1 = \omega_2$, and decreases as the distance between ω_1 and ω_2 becomes larger.

5 Endogenous Information

So far we have only discussed the cases where the signal process is exogenously determined and independent of agents' actions. This section we consider the case where an observed signal contains endogenous information.

An important theme of the literature on dispersed information is the role of the endogenous signal in coordinating beliefs and revealing information. Kasa (2000) and Pearlman and Sargent (2005) show that by observing prices in other industries, agents share the same beliefs. Walker (2007) and Rondina and Walker (2017) show that whether the price in the asset market reveals the state of the economy depends on whether the underlying shock follows a confounding process or not. However, most of the studies with endogenous information restrict their attention to the case in which the number of signals equals the number of shocks and agents observe the endogenous variable without noise to obtain analytic solution. In this section, we will analyse the role of endogenous information when there are more shocks than signals, and the endogenous variable cannot be observed perfectly.

5.1 Infinite state variables

In this subsection, we will provide an example to show that the finite-state representation result developed in previous sections will not hold when the signals contain endogenous variables. The model we use is a modification of the one used in Angeletos and La'O (2010). Assume that there is a continuum of agents and each agent uses the following independent value best response

$$y_{it} = \xi_{it} + \alpha \mathbb{E}_{it}[y_t]. \tag{5.1}$$

Agents observe their own fundamental ξ_{it} perfectly, which depends on both an aggregate component ξ_t and an idiosyncratic component ϵ_{it}

$$\xi_{it} = \xi_t + \sigma_\epsilon \epsilon_{it}. \tag{5.2}$$

As usual, we assume that the aggregate component ξ_t follows an AR(1) process

$$\xi_t = \rho \xi_{t-1} + \sigma_\eta \eta_t. \tag{5.3}$$

In terms of the information set, agents receive two signals each period, $\mathbf{x}_{it} = \begin{bmatrix} x_{it}^1 & x_{it}^2 \end{bmatrix}$. Their own idiosyncratic fundamental ξ_{it} is an exogenous signal about the aggregate fundamental ξ_t. We denote $x_{it}^1 = \xi_{it}$. In addition, agents also observe a second signal, which is the aggregate action y_t with an idiosyncratic noise u_{it}

$$x_{it}^2 = y_t + \sigma_u u_{it}. \tag{5.4}$$

The aggregate action y_t is endogenously determined by all the individual agents' choices, while at the same time, it enters agents' information set and affects the optimal choice of agents. In this case, we find it is more convenient to define the equilibrium with the innovation form.

Definition 5.1. *The equilibrium is a causal-stationary policy rule* $\boldsymbol{\phi}(L) = \{\phi_\epsilon(L), \phi_u(L), \phi_\eta(L)\}$, *and the law of motion* $\varphi(L)$ *for aggregate* y_t*, such that*

1. *Agent* i*'s information set* $\Omega_{it} = \left\{ x_{it}^1,\ x_{it}^2,\ x_{it-1}^1,\ x_{it-1}^2,\ \ldots \right\}$ *is determined by*

$$\begin{aligned} x_{it}^1 &= \xi_t + \sigma_\epsilon \epsilon_{it}, & (5.5) \\ x_{it}^2 &= \varphi(L)\eta_t + \sigma_u u_{it}, & (5.6) \end{aligned}$$

2. *Individual rationality*

$$y_{it} = \xi_{it} + \alpha \mathbb{E}_{it}[y_t] = \phi_\epsilon(L)\epsilon_{it} + \phi_u(L)u_{it} + \phi_\eta(L)\eta_t. \tag{5.7}$$

3. *Aggregate consistency*

$$y_t = \varphi(L)\eta_t = \int y_{it} = \phi_\eta(L)\eta_t. \tag{5.8}$$

The equilibrium with endogenous information involves two fixed points, the individual rationality and the consistency in terms of the signal process. Our preferred interpretation of the equilibrium conditions is the following. Whether or not the information set contains endogenous aggregate variables, each individual agent will behave competitively and take the signal process as exogenously given. For any $\varphi(L)$ in the signal process (5.6), one can solve the exogenous information equilibrium that satisfies conditional 1 and 2. If $\varphi(L)$ follows a finite ARMA process, the exogenous information equilibrium permits a finite-state representation. Condition 3 further selects a particular process $\varphi(L)$ that is simultaneously consistent with agents' choices. It requires that agents perceived law of motion of the aggregate y_t is the same as the actual law of motion of the aggregate y_t. This can be viewed as an additional cross-equation restriction in the sense that agents perception is in line with the reality generated by their own action.

The following proposition guarantees that there exists an equilibrium with endogenous information.

Proposition 5.1. *If $\alpha \in (0,1)$, then there exists an equilibrium of the model in Definition 5.1.*

Proof. See Appendix A.14 for proof. □

This proposition only proves the existence of the equilibrium, but it is silent on whether the agents need to keep track of infinite number of state variables or not. With exogenous information, we have shown that the equilibrium always permits a finite-state representation provided that the signals follow a finite ARMA process. In contrast, the following theorem shows that with endogenous information, the aggregate y_t does not follow a finite ARMA process. As a result, the equilibrium cannot have a finite-state representation.

Theorem 2. *If $\alpha \in (0,1)$, the equilibrium of the model in Definition 5.1 does not have a finite-state representation.*

Proof. See Appendix A.15 for proof. □

The proof of this theorem shows that if assuming the perceived aggregate y_t follows a finite ARMA process, the implied actual aggregate y_t cannot be the same as the perceived aggregate y_t. With endogenous information, if we assume the perceived y_t follows an AR(1) process, the implied actual y_t follows an ARMA $(2,1)$ process. If we assume perceived y_t follows ARMA

$(2,1)$, the actual y_t will follow an ARMA $(4,2)$ process. Iterating this process, the aggregate y_t follows an infinite ARMA process in the limit.

This is a somewhat surprising result. Kasa (2000) and Pearlman and Sargent (2005) show that in the Townsend (1983) model, there is actually no infinite regress problem and the equilibrium permits a finite-state representation. Similarly, in Rondina and Walker (2017) and Acharya (2013), the equilibrium policy rule has a finite-state representation as well. Pearlman and Sargent (2005) suspects that to resuscitate the infinite regress problem, there should be more shocks than signals. Theorem 2 proves that in our model with endogenous information, agents need to keep track of infinite state variables in equilibrium. Chari (1979) proved a similar impossibility theorem for a particular univariate case, and we prove this theorem in a multivariate system.

The reason for the infinite state variables, however, is not only due to the infinite higher order beliefs. When the signals follow an exogenous ARMA process, the infinite regress problem does exist but the equilibrium rule always has a finite-state representation. With endogenous information, each individual still treats the signal process as exogenous. If the perceived law of motion for y_t is a finite ARMA process, we return to the case covered by Theorem 1: each individual needs to solve the infinite regress problem, but the number of state variables is finite. With endogenous information, what complicates the issue is that the signal process itself cannot be represented as a finite ARMA process, but this is independent of the infinite regress problem faced by each individual.

Compared with the literature, the equilibrium policy rule in Kasa (2000), Rondina and Walker (2017) and Acharya (2013) all follows an ARMA process, even though the signals contain endogenous information. The key difference is that they assume the number of signals equals the number shocks, i.e., the signals $\mathbf{x}_t = \mathbf{M}(L)\mathbf{s}_t$ with $\mathbf{M}(L)$ being a square matrix. In this case, one can use the Blaschke matrix to obtain the Wold representation without knowing the exact signal process. The cost of this assumption is that the signal process cannot be too complicated. In Acharya (2013) or Kasa (2000), the endogenous variable that has an information role is observed without noise, and the forecast error is transitory. In our model, because there are more shocks than signals, agents can never infer the shocks perfectly, and the forecast error is persistent.

5.2 Computation

The infinite-state result is theoretically interesting, but it excludes the possibility of obtaining the exact solution. Here we provide a tractable algorithm that can approximate the true solution. The idea is to use a low order ARMA process to approximate endogenous aggregate actions that enter the information set. This will enable us to use the Winer-Hopf prediction formula.

1. Assume that the information process is given by

$$\begin{aligned} x_{it}^1 &= \xi_t + \sigma_\epsilon \epsilon_{it}, \\ x_{it}^2 &= \varphi^{(0)}(L)\eta_t + \sigma_u u_{it} \end{aligned}$$

   where $\varphi^{(0)}(L)$ follows a finite ARMA process

$$\varphi^{(0)}(L) = \sigma_y \frac{\Pi_{k=0}^q (L - \theta_k)}{\Pi_{k=1}^p (1 - \rho_k L)}$$

2. Given the signal process, we can use the method in Section 3 to solve for the policy rule $\boldsymbol{\phi} = \{\phi_\epsilon(L), \phi_u(L), \phi_\eta(L)\}$ in the exogenous information equilibrium. The implied actual aggregate y_t follows

$$y_t = \phi_\eta(L)\eta_t.$$

3. Compute the difference between the perceived law of motion and actual law of motion, $\|\varphi^{(0)} - \phi_\eta\|$[9]. If $\|\varphi^{(0)} - \phi_\eta\|$ is larger than the tolerance level, set the new law of motion that enters the signal process as $\varphi^{(1)}(L) = \phi_\eta(L)$ and repeat 1 to 3.

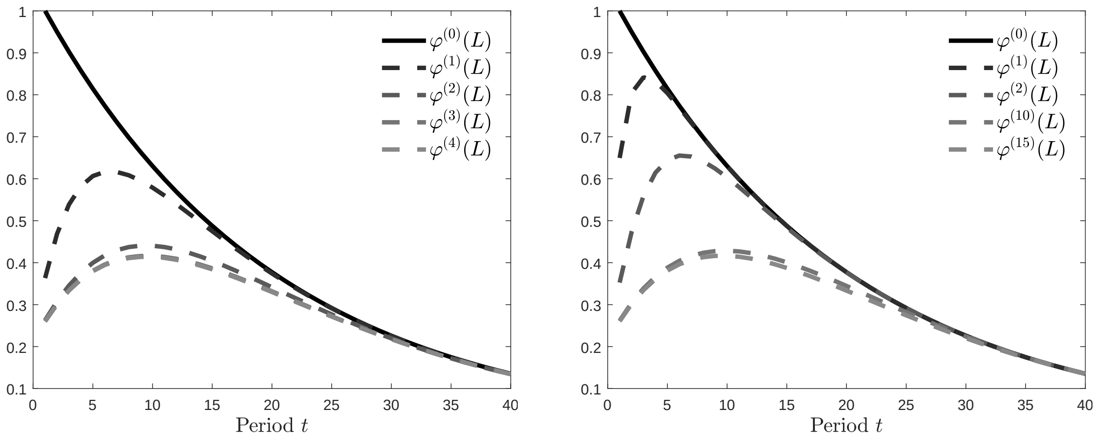
Crucially, in step 2, agents' perceived law of motion of y_t is $\phi_\eta(L)$ rather than $\varphi^{(0)}(L)$. $\varphi^{(0)}(L)$ only specifies the law of motion of the signal process. In this step, we still solve a fixed point problem with exogenous information, and agents' perceived law of motion and the actual law of motion of y_t are the same. Alternatively, one can assume that the perceived law of motion of y_t is simply $\varphi^{(0)}(L)$. In this case, the perceived law of motion and the actual law of motion of y_t are different. We call this alternative strategy as naive strategy. We show that our procedure is much more efficient than this alternative strategy.

[9] It is natural to use the ℓ^2 norm.

Sargent (1991) also uses an ARMA process to approximate the signal process, but our method differs from his in an important way. In Sargent (1991), only the forecasts of future signals are pay-off relevant. Once the law of motion of the signal is specified, agents do not need to solve the signal extraction problem and there is no need to forecast the forecasts of others. In our case, although the signal process is given, agents still face the infinite regress problem. Step 2 in the algorithm makes sure that each individual always performs their optimal prediction.

Compared with Nimark (2017), our method has the following advantages. The first advantage is that our method requires fewer state variables. Nirmark's method needs to keep track of a relatively large number of higher order beliefs to accurately approximate the policy rule. In our numerical example, it requires to keep track of the higher order beliefs up to order 30 to achieve the same accuracy as our ARMA (4,2) approximation. The second advantage is that our method is applicable in more general environments. Nirmark's method relies on the assumption that the best response of agents' action or aggregate law of motion can be represented by a weighted average of higher order beliefs. This representation can be difficult to obtain when the best reponse is complicated (see the quantitative model in Huo and Takayama (2017) for example). Our method does not rely on this assumption. The benefit of using Nirmark's method is that one can apply the Kalman filter for the inference problem and stay in the time domain without switching to the Wiener filter. In addition, Nirmark's method has a natural behavioral interpretation with bounded rationality.

Example We solved the example in section 5.1 numerically. We assume the initial guess of $\varphi^{(0)}(L) = \frac{1}{1-\alpha}\frac{1}{1-\rho L}$, which is consistent with the perfect information solution. The left panel in Figure 5 shows the iterations using the procedure discussed in this section. As can be seen, after two iterations, the law of motion of y_t almost converges. In contrast, as shown in the right panel of Figure 5, if one uses the alternative naive strategy by assuming $y_t = \varphi^{(0)}(L)\eta_t$, it takes more than 15 iterations to achieve the similar accuracy. Our preferred strategy requires a much smaller number of state variables and is more efficient. Given the existence of the equilibrium, this method can easily extend to other more complicated environments when there does not exist a finite-state representation.



(A) Compute exogenous info equilibrium

(B) Alternative naive strategy

FIGURE 5: IRF in Endogenous Information Equilibrium

Parameters: $\rho = 0.95, \alpha = 0.8, \sigma_\eta = 1, \sigma_u = \sigma_\epsilon = 4.$

6 Conclusion

In this paper, we have shown how to solve general rational expectations models with higher order beliefs. When the signal follows an ARMA process, we prove that the policy rule always admits a finite-state representation. It turns out the infinite regress problem does not require infinite state variables, because the total effects of the higher order beliefs can be summarized by a small set of variables. We provide a procedure that gives an explicit solution formula. The key of our method is to apply the Kalman filter to obtain the Wold representation of the signal process, and then use the Wiener filter to solve the inference problems. We also prove that when the signal process contains endogenous information, the equilibrium policy rule may not have a finite-state representation, which is in some sense the 'true' infinite regress problem. This is due to the fact that cross-equation restriction imposes an additional equilibrium condition that the perceive law of motion of an endogenous variable has to be the same as the law of motion that is generated by agents' actions. We provide a tractable algorithm that can approximate the true solution accurately with a small number of state variables. Various applications are easily solved by our method. We expect that the method we develop in this paper can be applied in a much broader class of models, especially in the areas of macroeconomics and financial economics with dispersed information.

References

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

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.

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

Baxter, B., L. Graham, and S. Wright (2011): "Invertible and Non-invertible Information Sets in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 35(3), 295–311.

Chari, V. (1979): "Linear Stochastic Models with Decentralized Information," Discussion Papers 458, Northwestern University.

Futia, C. A. (1981): "Rational Expectations in Stationary Linear Models," *Econometrica*, 49(1), 171–92.

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

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

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 (2017): "Infinite Higher Order Beliefs and First Order Beliefs: An Equivalence Result," Working paper, Yale University.

Huo, Z., and N. Takayama (2017): "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," Forthcoming, *The Review of Economic Studies*.

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

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," Economics Working Papers 1118.

Pearlman, J. G., and T. J. Sargent (2005): "Knowing the Forecasts of Others," *Review of Economic Dynamics*, 8(2), 480–497.

Perez, D., and A. Drenik (2015): "Price Setting Under Uncertainty About Inflation," Discussion paper.

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

Rondina, G., and T. B. Walker (2017): "Confounding Dynamics," Working paper.

Sargent, T. J. (1987): *Macroeconomic Theory*. Academic Press, New York, 2 edn.

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

Singleton, K. J. (1987): *Asset Prices in a Time-Series Model with Disparately Informed, Competitive Traders*. New Approaches to Monetary Economics, Eds. WA Burnett and KJ Singleton, Cambridge University Press.

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

Walker, T. B. (2007): "How equilibrium prices reveal information in a time series model with disparately informed, competitive traders," *Journal of Economic Theory*, 137(1), 512–537.

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

Appendix

A Proof of Theorems and Propositions

A.1 Riesz-Fisher Theorem

Theorem (Riesz-Fisher)**.** *Let $\{c_\tau\}$ be a square-summable sequence of complex numbers for which $\sum_{\tau=-\infty}^{\infty} |c_\tau|^2 < \infty$. Then there exists a complex-valued function $g(z)$, defined at least on the unit circle in the complex plane such that*

$$g(z) = \sum_{\tau=-\infty}^{\infty} c_\tau z^\tau,$$

where the infinite series converges in the mean square sense that

$$\lim_{n\to\infty} \oint \left| \sum_{\tau=-n}^{n} c_\tau z^\tau - g(z) \right|^2 \frac{dz}{z} = 0$$

where the integral is a contour integral on the unit circle. The function $g(z)$ is square-integrable

$$\left| \frac{1}{2\pi i} \oint |g(z)|^2 \frac{dz}{z} \right| < \infty$$

The function $g(z)$ is called the z transform of the sequence $\{c_\tau\}$.

Conversely, given a square-integrable $g(z)$, there exists a square- summable sequence $\{c_\tau\}$ where

$$c_\tau = \frac{1}{2\pi i} \oint g(z) z^{-\tau-1} dz.$$

Furthermore, suppose $\{c_\tau\}$ be a one-side square-summable sequence for which $\sum_{\tau=0}^{\infty} |c_\tau|^2 < \infty$. Then there exists an analytic function $g(z)$ on the open unit disk such that

$$g(z) = \sum_{\tau=0}^{\infty} c_\tau z^\tau.$$

Conversely, given an analytic function on the unit disk, there exists a one-side square-summable sequence $\{c_\tau\}$ where

$$c_\tau = \frac{1}{2\pi i} \oint g(z) z^{-\tau-1} dz.$$

Proof. The proof of this theorem is referred to Sargent (1987) and Kasa (2000). □

A.2 Proof of Lemma 3.1

Proof. There can be many different state-space representations and we only give one of them here, which is sufficient to prove the claim. Hamilton (1994) shows how to represent a univariate ARMA process in state space, and what we construct below is a natural extension to the multivariate case.

Without loss of generality, we normalize $\delta_{ij0} = 1$. Let $u_{ij} = \max\{p_{ij}, q_{ij} + 1\}$, and $u = \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij}$. $\mathbf{F}_{u\times u}$ is constructed in the following way

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{12} & \mathbf{0} & \dots & \dots & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \dots & \dots & \dots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{F}_{1m} & \dots & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \dots & \mathbf{F}_{nm-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{F}_{nm} \end{bmatrix}. \tag{A.1}$$

The element $\mathbf{F}_{ij}$ in $\mathbf{F}$ is a $u_{ij} \times u_{ij}$ matrix

$$\mathbf{F}_{ij} = \begin{bmatrix} \delta_{ij1} & \delta_{ij2} & \dots & \delta_{iju_{ij}-1} & \delta_{iju_{ij}} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

$\mathbf{Q}$ is a $u \times m$ matrix with the following form

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} \\ \mathbf{Q}_{12} \\ \vdots \\ \mathbf{Q}_{1m} \\ \vdots \\ \mathbf{Q}_{nm-1} \\ \mathbf{Q}_{nm} \end{bmatrix}, \tag{A.2}$$

and each element $\mathbf{Q}_{ij}$ in $\mathbf{Q}$ is a $u_{ij} \times m$ matrix

$$\mathbf{Q}_{ij} = \begin{bmatrix} 0 & \dots & 1 & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}, \tag{A.3}$$

where 1 is at the j-th column.

$\mathbf{H}$ is a $n \times u$ matrix with the following form

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \dots & \mathbf{H}_{1m} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{H}_{21} & \dots & \mathbf{H}_{2m} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots & \ddots & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{H}_{n1} & \dots & \mathbf{H}_{nm} \end{bmatrix} \tag{A.4}$$

The element $\mathbf{H}_{ij}$ in $\mathbf{H}$ is a $1 \times u_{ij}$ matrix

$$\mathbf{H}_{ij} = \begin{bmatrix} \alpha_{ij0} & \alpha_{ij1} & \dots & \alpha_{iju_{ij-1}} \end{bmatrix}.$$

Let $\mathbf{Z}_t$ follows

$$\mathbf{Z}_t = \mathbf{F}\mathbf{Z}_{t-1} + \mathbf{Q}\mathbf{s}_t.$$

We have

$$\mathbf{x}_t = \mathbf{M}(L)\mathbf{s}_t = \mathbf{H}\mathbf{Z}_t$$

To show that the eigenvalues of $\mathbf{F}$ lie inside the unit circle, we can iterate the $\mathbf{Z}_t$ to obtain

$$\mathbf{Z}_t = \sum_{j=0}^{\infty} \mathbf{F}^j L^j \mathbf{Q}\mathbf{s}_t = (\mathbf{I} - \mathbf{F}L)^{-1}\mathbf{Q}\mathbf{s}_t$$

If the eigenvalues of $\mathbf{F}$ lies outside the unit circle, then $\mathbf{Z}_t$ is not co-variance stationary, which contradicts the assumption that $\mathbf{x}_t$ is co-variance stationary. □

A.3 Proof of Wiener-Hopf Theorem

Proof. A formal proof can be found in Whittle (1983). Here we provide a sketch of the proof.

Suppose the prediction is based on all the realization of the signals $\mathbf{x}^\infty$ instead of $\mathbf{x}^t$. The optimal linear prediction of f_t is

$$\mathbb{E}[f_t|\mathbf{x}^\infty] = \boldsymbol{\rho}_{fx}(L)\boldsymbol{\rho}_{xx}(L)^{-1}\mathbf{x}_t.$$

This formula is the same as that in OLS regression. $\boldsymbol{\rho}_{fx}$ measures the correlation between f_t and $\mathbf{x}_t$, adjusted by $\boldsymbol{\rho}_{xx}$. Given the fundamental representation

$$\mathbf{x}_t = \mathbf{B}(L)\mathbf{w}_t,$$

the prediction is equivalent to the prediction conditional on $\mathbf{w}^\infty$ and the prediction formula can be written as

$$\begin{aligned}\mathbb{E}[f_t|\mathbf{x}^\infty] =&\mathbb{E}[f_t|\mathbf{w}^\infty]\\ =&\boldsymbol{\rho}_{fx}(L)\boldsymbol{\rho}_{xx}(L)^{-1}\mathbf{x}_t,\\ =&\boldsymbol{\rho}_{fx}(L)\mathbf{B}'(L^{-1})^{-1}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{B}(L)w_t,\\ =&\boldsymbol{\rho}_{fx}(L)\mathbf{B}'(L^{-1})^{-1}\mathbf{V}^{-1}\mathbf{w}_t.\end{aligned}$$

Now imagine the prediction is conditional on only current and past signals $\mathbf{x}^t$, which is equivalent to conditional on $\mathbf{w}^t$. Since $\mathbf{w}_t$ is serially uncorrelated, the best forecast of $\mathbf{w}_k$ for $k > t$ is zero. Note that $\boldsymbol{\rho}_{fx}(L)\mathbf{B}'(L^{-1})^{-1}$ contains negative powers of L and the best forecast of $\mathbf{w}_k$ for $k > t$ is zero, the optimal prediction for f_t is simply

$$\begin{aligned}\mathbb{E}[f_t|\mathbf{x}^t] =&\mathbb{E}[f_t|\mathbf{w}^t]\\ =&[\boldsymbol{\rho}_{fx}(L)\mathbf{B}'(L^{-1})^{-1}]_+\mathbf{V}^{-1}\mathbf{w}_t,\\ =&[\boldsymbol{\rho}_{fx}(L)\mathbf{B}'(L^{-1})^{-1}]_+\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{x}_t.\end{aligned}$$

Recall that $\mathbf{B}(L)$ is invertible, so $\mathbf{B}(L)^{-1}$ contains only positive powers of L. □

A.4 Proof of Proposition 3.1

Proof. From Canonical Factorization 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

$$\mathbf{Z}_t = (\mathbf{I} - \mathbf{F}L)^{-1}\,\mathbf{Q}\mathbf{s}_t, \text{ and, } \mathbf{x}_t = \mathbf{H}\mathbf{Z}_t,$$

which implies

$$\mathbf{x}_t = \mathbf{H}\,(\mathbf{I} - \mathbf{F}L)^{-1}\,\mathbf{Q}\mathbf{s}_t.$$

By definition, it follows that

$$\mathbf{M}(L) = \mathbf{H}\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q}.$$

Hence,

$$\begin{aligned}
\mathbf{B}(L)^{-1}\mathbf{M}(L) &= \Big(\mathbf{I} - \mathbf{H}\left[\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right]^{-1}\mathbf{FK}L\Big)\,\mathbf{H}\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q} \\
&= \mathbf{H}\Big(\mathbf{I} - \left[\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right]^{-1}\mathbf{FKH}L\Big)\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q} \\
&= \mathbf{H}\Big(\mathbf{I} - \left[\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right]^{-1}\left(\left(\mathbf{I} - \mathbf{F}L\right) + \mathbf{FKH}L - \left(\mathbf{I} - \mathbf{F}L\right)\right)\Big)\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q} \\
&= \mathbf{H}\Big(\mathbf{I} - \mathbf{I} + \left[\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right]^{-1}\left(\mathbf{I} - \mathbf{F}L\right)\Big)\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q} \\
&= \mathbf{H}\left[\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right]^{-1}\left(\mathbf{I} - \mathbf{F}L\right)\left(\mathbf{I} - \mathbf{F}L\right)^{-1}\mathbf{Q} \\
&= \frac{\mathbf{H}\,\mathrm{Adj}\left(\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right)\mathbf{Q}}{\det\left(\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right)} \\
&= \frac{\mathbf{H}\,\mathrm{Adj}\left(\mathbf{I} - \left(\mathbf{F} - \mathbf{FKH}\right)L\right)\mathbf{Q}}{\prod_{k=1}^{u}\left(1 - \lambda_k L\right)},
\end{aligned}$$

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

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

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

$$\begin{aligned}
\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'(L^{-1})^{-1} &= \frac{\mathbf{Q}'\,\mathrm{Adj}\left(\mathbf{I} - \left(\mathbf{F}' - \mathbf{H}'\mathbf{K}'\mathbf{F}'\right)L^{-1}\right)\mathbf{H}'}{\prod_{k=1}^{u}\left(1 - \lambda_k L^{-1}\right)} \\
&= \frac{\mathbf{Q}'\,\mathrm{Adj}\left(\mathbf{I}L - \left(\mathbf{F}' - \mathbf{H}'\mathbf{K}'\mathbf{F}'\right)\right)\mathbf{H}'L}{L^{v-u}\prod_{k=1}^{u}\left(L - \lambda_k\right)}, \\
&\equiv \frac{\mathbf{G}(L)}{\prod_{k=1}^{u}\left(L - \lambda_k\right)}
\end{aligned}$$

By the Wiener-Hopf Theorem, the prediction formula is

$$\mathbb{E}\left[f_t \mid \mathbf{x}^t\right] = \left[\boldsymbol{\psi}(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_{+}\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{x}_t$$

and by assumption and previous derivation,

$$\begin{aligned}
\mathbf{M}'\left(L^{-1}\right)\mathbf{B}'(L^{-1})^{-1} &= \frac{\mathbf{G}(L)}{\prod_{k=1}^{u}\left(L - \lambda_k\right)}, \\
\boldsymbol{\psi}(L) &= \frac{\mathbf{a}(L)}{\prod_{\tau=1}^{d}(L - \beta_\tau)}.
\end{aligned}$$

The following result is useful for analysing the plus operator. Suppose $g(z)$ is a rational function of z that does not contains negative powers of z in expansion and $|\xi_k| < 1$, then

$$\left[\frac{g(z)}{(z-\xi_1)\cdots(z-\xi_\ell)}\right]_+ = \frac{g(z)}{(z-\xi_1)\cdots(z-\xi_\ell)} - \sum_{k=1}^{\ell} \frac{g(\xi_k)}{(z-\xi_k)\Pi_{\tau\neq k}(\xi_k-\xi_\tau)}$$

Based on this result, we have

$$\begin{aligned}\left[\boldsymbol{\psi}(L)\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_+ \quad = \quad & \frac{\mathbf{a}(L)}{\prod_{\tau=1}^{d}(L-\beta_\tau)}\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} \\ & -\sum_{k=1}^{u} \frac{\mathbf{a}(\lambda_k)\mathbf{G}(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} \\ & -\sum_{k=1}^{d} \frac{\mathbf{a}(\beta_k)\mathbf{G}(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\end{aligned}$$

Note that

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

The result then follows straightforwardly. □

A.5 Proof of Proposition 3.2

Proof. By Proposition 3.1, the forecast about $\mathbb{E}_t[\mathbf{P}(L)\mathbf{y}_t]$ can be written as

$$\mathbb{E}_t[\mathbf{P}(L)\mathbf{y}_t] = \mathbb{E}_t\left[\mathbf{P}(L)\begin{bmatrix} y_{1t} \\ \vdots \\ y_{rt}\end{bmatrix}\right] = \mathbb{E}_t\left[\mathbf{P}(L)\begin{bmatrix}\boldsymbol{\phi}(L)'\mathbf{A}_1\mathbf{M}(L)\mathbf{s}_t \\ \vdots \\ \boldsymbol{\phi}(L)'\mathbf{A}_r\mathbf{M}(L)\mathbf{s}_t\end{bmatrix}\right] = \mathbb{E}_t\left[\begin{bmatrix}\sum_{i=1}^{r} P_{1i}(L)\boldsymbol{\phi}(L)'\mathbf{A}_i\mathbf{M}(L)\mathbf{s}_t \\ \vdots \\ \sum_{i=1}^{r} P_{ri}(L)\boldsymbol{\phi}(L)'\mathbf{A}_i\mathbf{M}(L)\mathbf{s}_t\end{bmatrix}\right]$$

By Proposition 3.1, it follows that the forecast about $P_{ji}(L)y_{it}$ is given by

$$\begin{aligned}\mathbb{E}_t[P_{ji}(L)\boldsymbol{\phi}(L)'\mathbf{A}_i\mathbf{M}(L)\mathbf{s}_t] =& P_{ji}(L)\boldsymbol{\phi}(L)'\mathbf{A}_i\mathbf{M}(L)\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\mathbf{x}_t \\ & -\sum_{k=1}^{u} \frac{P_{ji}(\lambda_k)\boldsymbol{\phi}(\lambda_k)'\mathbf{A}_i\mathbf{M}(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(L)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)}\mathbf{x}_t \\ & -\sum_{k=1}^{d} \frac{P_{ji}(\beta_k)\boldsymbol{\phi}(\beta_k)'\mathbf{A}_i\mathbf{M}(\beta_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(L)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\mathbf{x}_t\end{aligned}$$

The forecast about $\mathbf{Q}(L)\widetilde{\mathbf{y}}_t$ and $\mathbf{R}(L)\mathbf{s}_t$ can be obtained in a similar way. Collecting all terms and using the fact that $\mathbf{M}(L)\mathbf{M}'(L^{-1}) = \boldsymbol{\rho}_{xx}(L)$, the system of equations becomes

$$\sum_{i=1}^{r} \begin{bmatrix} \boldsymbol{\phi}(L)'\mathbf{A}_i(P_{1,i}(L)\mathbf{I}_n + Q_{1,i}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}) \\ \vdots \\ \boldsymbol{\phi}(L)'\mathbf{A}_i(P_{r,i}(L)\mathbf{I}_n + Q_{r,i}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}) \end{bmatrix} \mathbf{x}_t =$$

$$\begin{bmatrix} \Big(-\mathbf{R}_1(z)\mathbf{M}'(z^{-1})\boldsymbol{\rho}_{xx}(z)^{-1} + \sum_{k=1}^{u} \frac{\widehat{\mathbf{R}}_1(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{r} \frac{\widehat{\mathbf{R}}_1(\lambda_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{u} \frac{(\widehat{P}_{1,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k) + \widehat{Q}_{1,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)\boldsymbol{\Lambda})\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{d} \frac{(\widehat{P}_{1,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k) + \widehat{Q}_{1,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)\boldsymbol{\Lambda})\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \Big)\mathbf{x}_t \\ \vdots \\ \Big(-\mathbf{R}_r(z)\mathbf{M}'(z^{-1})\boldsymbol{\rho}_{xx}(z)^{-1} + \sum_{k=1}^{u} \frac{\widehat{\mathbf{R}}_r(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{r} \frac{\widehat{\mathbf{R}}_r(\lambda_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{u} \frac{(\widehat{P}_{r,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k) + \widehat{Q}_{r,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)\boldsymbol{\Lambda})\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{d} \frac{(\widehat{P}_{r,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k) + \widehat{Q}_{r,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)\boldsymbol{\Lambda})\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \Big)\mathbf{x}_t \end{bmatrix}$$

This has to be true for all the possible realizations of $\{\mathbf{x}_t\}$. Note that

$$\begin{aligned} &\sum_{i=1}^{r} \begin{bmatrix} \big(\boldsymbol{\phi}(L)'\mathbf{A}_i(P_{1,i}(L)\mathbf{I}_n + Q_{1,i}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1})\big)' \\ \vdots \\ \big(\boldsymbol{\phi}(L)'\mathbf{A}_i(P_{r,i}(L)\mathbf{I}_n + Q_{r,i}(L)\mathbf{M}(L)\boldsymbol{\Lambda}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1})\big)' \end{bmatrix} \\ =&\sum_{i=1}^{r}(\mathbf{P}_i(L)\otimes\mathbf{I}_n)(\mathbf{e}_i'\otimes\mathbf{I}_n)\boldsymbol{\phi}(L) + \sum_{i=1}^{r}(\mathbf{I}_r\otimes\boldsymbol{\rho}_{xx}'(L)^{-1}\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L))(\mathbf{Q}_i(L)\otimes\mathbf{I}_n)(\mathbf{e}_i'\otimes\mathbf{I}_n)\boldsymbol{\phi}(L) \\ =&\sum_{i=1}^{r}(\mathbf{P}_i(L)\mathbf{e}_i'(L)\otimes\mathbf{I}_n)\boldsymbol{\phi}(L) + (\mathbf{I}_r\otimes\boldsymbol{\rho}_{xx}'(L)^{-1}\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L))\left(\sum_{i=1}^{r}\mathbf{Q}_i(L)\mathbf{e}_i'\otimes\mathbf{I}_n\right)\boldsymbol{\phi}(L) \\ =&(\mathbf{P}(L)\otimes\mathbf{I}_n)\boldsymbol{\phi}(L) + (\mathbf{Q}(L)\otimes\boldsymbol{\rho}_{xx}'(L)^{-1}\mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L))\boldsymbol{\phi}(L) \end{aligned}$$

$$\begin{bmatrix} \Big(-\mathbf{R}_1(z)\mathbf{M}'(z^{-1})\boldsymbol{\rho}_{xx}(z)^{-1} + \sum_{k=1}^{u} \frac{\widehat{\mathbf{R}}_1(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{r} \frac{\widehat{\mathbf{R}}_1(\lambda_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{u} \frac{(\widehat{P}_{1,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)+\widehat{Q}_{1,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)\boldsymbol{\Lambda})\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{d} \frac{(\widehat{P}_{1,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)+\widehat{Q}_{1,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)\boldsymbol{\Lambda})\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \Big)' \\ \vdots \\ \Big(-\mathbf{R}_r(z)\mathbf{M}'(z^{-1})\boldsymbol{\rho}_{xx}(z)^{-1} + \sum_{k=1}^{u} \frac{\widehat{\mathbf{R}}_r(\lambda_k)\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{r} \frac{\widehat{\mathbf{R}}_r(\lambda_k)\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{u} \frac{(\widehat{P}_{r,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)+\widehat{Q}_{r,i}(\lambda_k)\boldsymbol{\phi}'(\lambda_k)\mathbf{A}_i\mathbf{M}(\lambda_k)\boldsymbol{\Lambda})\mathbf{G}(\lambda_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} \\ + \sum_{i=1}^{r}\sum_{k=1}^{d} \frac{(\widehat{P}_{r,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)+\widehat{Q}_{r,i}(\beta_k)\boldsymbol{\phi}'(\beta_k)\mathbf{A}_i\mathbf{M}(\beta_k)\boldsymbol{\Lambda})\mathbf{G}(\beta_k)\mathbf{V}^{-1}\mathbf{B}(z)^{-1}}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \Big)' \end{bmatrix}$$

$$\begin{aligned} &= -(\mathbf{I}_r \otimes \boldsymbol{\rho}'_{xx}(L)^{-1}\mathbf{M}(L^{-1}))\mathbf{vec}(\mathbf{R}'(L)) \\ &+ \sum_{k=1}^{u} \frac{(\mathbf{I}_r \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\lambda_k))\mathbf{vec}(\widehat{\mathbf{R}}'(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{d} \frac{(\mathbf{I}_r \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\beta_k))\mathbf{vec}(\widehat{\mathbf{R}}'(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ &+ \sum_{k=1}^{u} \frac{(\widehat{\mathbf{P}}(\lambda_k) \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\lambda_k)\mathbf{M}'(\lambda_k))\boldsymbol{\phi}(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{d} \frac{(\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k))\boldsymbol{\phi}(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ &+ \sum_{k=1}^{u} \frac{(\widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\lambda_k)\boldsymbol{\Lambda}\mathbf{M}'(\lambda_k))\boldsymbol{\phi}(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{d} \frac{(\widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\beta_k)\boldsymbol{\Lambda}\mathbf{M}'(\beta_k))\boldsymbol{\phi}(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \end{aligned}$$

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

$$\begin{aligned} &\Big(\mathbf{P}(L) \otimes \boldsymbol{\rho}'_{xx}(L) + \mathbf{Q}(L) \otimes \mathbf{M}(L^{-1})\boldsymbol{\Lambda}\mathbf{M}'(L)\Big)\boldsymbol{\phi}(L) \\ &= -\mathbf{vec}(\mathbf{M}(L^{-1})\mathbf{R}'(L)) + \sum_{k=1}^{u} \frac{\mathbf{vec}(\mathbf{B}(L^{-1})\mathbf{G}'(\lambda_k)\widehat{\mathbf{R}}'(\lambda_k))}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} + \sum_{k=1}^{d} \frac{\mathbf{vec}(\mathbf{B}(L^{-1})\mathbf{G}'(\beta_k)\widehat{\mathbf{R}}'(\beta_k))}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} \\ &+ (\mathbf{I}_r \otimes \mathbf{B}(L^{-1}))\sum_{k=1}^{u} \frac{\widehat{\mathbf{P}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{M}'(\lambda_k) + \widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\boldsymbol{\Lambda}\mathbf{M}'(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)}\boldsymbol{\phi}(\lambda_k) \\ &+ (\mathbf{I}_r \otimes \mathbf{B}(L^{-1}))\sum_{k=1}^{d} \frac{\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k) + \widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\boldsymbol{\Lambda}\mathbf{M}'(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\boldsymbol{\phi}(\beta_k) \end{aligned}$$

By Riesz-Fischer Theorem, it is equivalent to the following system of functional equations

$$\mathbf{T}(z)\boldsymbol{\phi}(z) = \mathbf{D}\Big[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=0}^{d}\Big]$$

By the Riesz-Fisher Theorem, there exists $\phi(L)$ that solves model (3.4) if and only if there exists a vector analytic function $\phi(z)$ that solves equations (3.17). □

A.6 Proof of Lemma 3.2

Proof. Consider the part related to $\boldsymbol{\phi}(\lambda_k)$ for example. There exist matrices $\mathbf{\Phi}_{k,1}$ and $\mathbf{\Phi}_{k,2}$ such that

$$\frac{\widehat{\mathbf{P}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{M}'(\lambda_k) + \widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{\Lambda}\mathbf{M}'(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)} = \frac{1}{L-\lambda_k}\mathbf{\Phi}_{k,1}\mathbf{\Phi}_{k,2}$$

where $\mathbf{\Phi}_{k,1}$ is with full column rank and $\mathbf{\Phi}_{k,2}$ is with full row rank. The rank of $\mathbf{\Phi}_{k,1}$ and $\mathbf{\Phi}_{k,2}$ are the same as the rank of $\widehat{\mathbf{P}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{M}'(\lambda_k) + \widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{\Lambda}\mathbf{M}'(\lambda_k)$.

Similarly, for the part related to $\boldsymbol{\phi}(\beta_k)$, there exist matrices $\mathbf{\Psi}_{k,1}$ and $\mathbf{\Psi}_{k,2}$ such that

$$\frac{\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k) + \widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{\Lambda}\mathbf{M}'(\beta_k)}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)} = \frac{1}{L-\beta_k}\mathbf{\Psi}_{k,1}\mathbf{\Psi}_{k,2}$$

where $\mathbf{\Psi}_{k,1}$ is with full column rank and $\mathbf{\Psi}_{k,2}$ is with full row rank.

Now we can rewrite $\mathbf{D}\Big[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=0}^{d}\Big]$ as

$$\mathbf{D}\Big[z, \{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=0}^{d}\Big] = \mathbf{D}_1(z)\boldsymbol{\psi} + \mathbf{D}_2(z)$$

where

$$\begin{aligned}
\mathbf{D}_1(z) &= (\mathbf{I}_r \otimes \mathbf{B}(z^{-1}))\begin{bmatrix}\frac{\mathbf{\Phi}_{1,1}}{z-\lambda_1} & \dots & \frac{\mathbf{\Phi}_{u,1}}{z-\lambda_u} & \frac{\mathbf{\Psi}_{1,1}}{z-\beta_1} & \dots & \frac{\mathbf{\Psi}_{d,1}}{z-\lambda_d}\end{bmatrix}\\
\mathbf{D}_2(z) &= -\mathbf{vec}(\mathbf{M}(L^{-1})\mathbf{R}'(L)) + \sum_{k=1}^{u}\frac{\mathbf{vec}(\mathbf{B}(L^{-1})\mathbf{G}'(\lambda_k)\widehat{\mathbf{R}}'(\lambda_k))}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)\prod_{\tau=1}^{d}(\lambda_k-\beta_\tau)}\\
&\quad+\sum_{k=1}^{d}\frac{\mathbf{vec}(\mathbf{B}(L^{-1})\mathbf{G}'(\beta_k)\widehat{\mathbf{R}}'(\beta_k))}{(L-\beta_k)\prod_{\tau=1}^{k}(\beta_k-\lambda_\tau)\prod_{\tau\neq k}^{d}(\beta_k-\beta_\tau)}\\
\boldsymbol{\psi} &= \begin{bmatrix}\mathbf{\Phi}_{1,2}\boldsymbol{\phi}(\lambda_1)\\ \vdots\\ \mathbf{\Phi}_{u,2}\boldsymbol{\phi}(\lambda_u)\\ \mathbf{\Psi}_{1,2}\boldsymbol{\phi}(\beta_1)\\ \vdots\\ \mathbf{\Psi}_{d,2}\boldsymbol{\phi}(\beta_d)\end{bmatrix}
\end{aligned}$$

Note that the dimension $\boldsymbol{\psi}$ is $N_1 \times 1$, where

$$N_1 = \sum_{k=1}^{u}\mathbf{rank}\Big(\widehat{\mathbf{Q}}(\lambda_k) \otimes \mathbf{G}'(\lambda_k)\mathbf{\Lambda}\mathbf{M}'(\lambda_k)\Big) + \sum_{k=1}^{d}\mathbf{rank}\Big(\widehat{\mathbf{P}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{M}'(\beta_k) + \widehat{\mathbf{Q}}(\beta_k) \otimes \mathbf{G}'(\beta_k)\mathbf{\Lambda}\mathbf{M}'(\beta_k)\Big).$$

Furthermore, $\mathbf{D}_1(z)$ is with full column rank. $\square$

A.7 Proof of Theorem 1

Proof. By Cramer's rule, the i-th element of $\boldsymbol{\phi}(z)$ that solves equation (3.17) is given by

$$\phi_i(z) = \frac{\det\left[\mathbf{T}_1(z) \quad \ldots \quad \mathbf{T}_{i-1}(z) \;\; \mathbf{D}_1(z)\boldsymbol{\psi} + \mathbf{D}_2(z) \;\; \mathbf{T}_{i+1}(z)\ldots \quad \ldots \quad \mathbf{T}_{rn}(z)\right]}{\det[\mathbf{T}(z)]}$$

By Assumption 3.2 and Proposition 3.1, the functions in $\mathbf{T}(z)$, $\mathbf{D}_1(z)$, and $\mathbf{D}_2(z)$ are all rational functions with finite degree. As a result, whether $\phi_i(z)$ is an analytic function or not is equivalent to whether $\phi_i(z)$ has poles within the unit circle or not.

The inside poles of $\phi_i(z)$ are either the inside roots of $\det[\mathbf{T}(z)]$, i.e., $\{\zeta_1, \ldots, \zeta_{N_2}\}$, or the inside poles of

$$\widehat{\phi}_i(z) \equiv \det\left[\mathbf{T}_1(z) \quad \ldots \quad \mathbf{T}_{i-1}(z) \;\; \mathbf{D}_1(z)\boldsymbol{\psi} + \mathbf{D}_2(z) \;\; \mathbf{T}_{i+1}(z)\ldots \quad \ldots \quad \mathbf{T}_{rn}(z)\right].$$

By construction, the only poles of $\widehat{\phi}_i(z)$ are $\{\lambda_k\}_{k=1}^{u}$ and $\{\beta_k\}_{k=1}^{d}$. However, $\{\lambda_k\}_{k=1}^{u}$ and $\{\beta_k\}_{k=1}^{d}$ cannot be poles of $\phi_i(z)$ because these poles are generated from the Wiener-Hopf prediction formula, and by Proposition 3.1, these poles are already eliminated by $\{\{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\}$. Therefore, we only need to consider the inside roots of $\det[\mathbf{T}(z)]$.

For any ζ_i, we have $\det[\mathbf{T}(\zeta_i)] = 0$ and there exits ℓ_i such that

$$\mathbf{T}_{\ell_i}(\zeta_i) = \sum_{k \neq \ell_i} \varphi_k^i \mathbf{T}_k(\zeta_i)$$

Suppose that

$$\det\left[\mathbf{T}_1(\zeta_i) \quad \ldots \quad \mathbf{T}_{\ell_i-1}(\zeta_i) \quad \mathbf{D}_1(\zeta_1)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_1) \quad \mathbf{T}_{\ell_i+1}(\zeta_i) \quad \ldots \quad \mathbf{T}_{rn}(\zeta_i)\right] = 0. \tag{A.5}$$

For any $j \neq \ell_i$, we have

$$\begin{aligned}
&\det\Big[\mathbf{T}_1(\zeta_i) \quad \ldots \quad \underbrace{\mathbf{D}_1(\zeta_i)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_i)}_{j\text{-th column}} \quad \ldots \quad \underbrace{\mathbf{T}_{\ell_i}(\zeta_i)}_{\ell_i\text{-th column}} \quad \ldots \quad \mathbf{T}_{rn}(\zeta_i)\Big] \\
= &\sum_{k \neq \ell_i} \det\Big[\mathbf{T}_1(\zeta_i) \quad \ldots \quad \underbrace{\mathbf{D}_1(\zeta_i)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_i)}_{j\text{-th column}} \quad \ldots \quad \underbrace{\varphi_k^i \mathbf{T}_k(\zeta_i)}_{\ell_i\text{-th column}} \quad \ldots \quad \mathbf{T}_{rn}(\zeta_i)\Big] \\
= &\; 0.
\end{aligned}$$

This implies that if equation (A.5) holds, then for all $j \in \{1, \ldots, rn\}$, ζ_i is the root of the determinant

$$\det \Big[\mathbf{T}_1(\zeta_i) \quad \ldots \underbrace{\mathbf{D}_1(\zeta_1)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_1)}_{j\text{-th column}} \quad \ldots \quad \mathbf{T}_{rn}(\zeta_i) \Big] = 0.$$

Consequently, ζ_i cannot be a pole of $\boldsymbol{\phi}(z)$.

Without loss of generality, assume that $\ell_i = 1$. To choose $\boldsymbol{\psi}$ to remove all poles of $\boldsymbol{\phi}(z)$, consider the following problem,

$$\mathbf{U}_1\boldsymbol{\psi} + \mathbf{U}_2 \equiv \begin{bmatrix} \det \Big[\mathbf{D}_1(\zeta_1)\boldsymbol{\psi} + \mathbf{D}_2(\zeta_1) & \mathbf{T}_2(\zeta_1) & \ldots & \mathbf{T}_{rn}(\zeta_1) \Big] \\ \vdots & \vdots & & \vdots \\ \det \Big[\mathbf{D}_1(\zeta_{N_2})\boldsymbol{\psi} + \mathbf{D}_2(\zeta_{N_2}) & \mathbf{T}_2(\zeta_{N_2}) & \ldots & \mathbf{T}_{rn}(\vartheta_{N_2}) \Big] \end{bmatrix}.$$

If there exists $\boldsymbol{\psi}$ such that

$$\mathbf{U}_1\boldsymbol{\psi} + \mathbf{U}_2 = \mathbf{0}, \tag{A.6}$$

then $\{\zeta_i\}_{i=1}^{N_2}$ are not poles of $\boldsymbol{\phi}(z)$.

1. If $N_1 < N_2$, then there are more equations than unknowns. There does not exist $\boldsymbol{\psi}$ such that equation (A.6) holds. As a result, there is no solution to (3.17).
2. If $N_1 = N_2 = \mathbf{rank}(\mathbf{U}_2)$, then there exists a unique $\boldsymbol{\psi}$ that solves (A.6). Therefore, $\{\zeta_i\}_{i=1}^{N_2}$ are not poles of $\boldsymbol{\phi}(z)$.
3. If $N_1 > N_2$ or $N_1 = N_2 > \mathbf{rank}(\mathbf{U}_2)$, there are infinite solutions to (A.6). As a result, there are infinite number of analytic functions $\boldsymbol{\phi}(z)$ that solves (3.17).

□

A.8 Proof of Proposition 3.3

Proof. By Assumption 3.2 and Proposition 3.1, the functions in $\mathbf{T}(z)$, $\mathbf{D}_1(z)$, and $\mathbf{D}_2(z)$ are all rational functions with finite degree. If there exists a solution to (3.4), $\boldsymbol{\phi}(z)$ is a rational function with finite degree. The inside poles of $\boldsymbol{\phi}(z)$ are removed by proper choices of $\{\{\boldsymbol{\phi}(\lambda_k)\}_{k=1}^{u}, \{\boldsymbol{\phi}(\beta_k)\}_{k=1}^{d}\}$, and the remaining poles can only be the outside roots of $\det[\mathbf{T}(z)]$, which are $\{\vartheta_1^{-1}, \ldots, \vartheta_{N_3}^{-1}\}$.

Denote $\widetilde{\mathbf{M}}(L) \equiv \mathbf{h}(L)$, $\widetilde{\mathbf{x}}_t \equiv \mathbf{y}_t$, and $\widetilde{\mathbf{s}}_t \equiv \mathbf{x}_t$. By Lemma 3.1, there exists a state space representation of $\mathbf{y}_t$, which is given by

$$\mathbf{z}_t = \widetilde{\mathbf{F}}\mathbf{z}_{t-1} + \widetilde{\mathbf{Q}}\mathbf{x}_t \tag{A.7}$$

$$\mathbf{y}_t = \widetilde{\mathbf{H}}\mathbf{z}_t \tag{A.8}$$

where $\widetilde{\mathbf{F}}, \widetilde{\mathbf{Q}}$ and $\widetilde{\mathbf{H}}$ can be constructed according to (A.1) to (A.4) respectively. Define

$$\mathbf{\Gamma}_x = \widetilde{\mathbf{Q}}, \quad \mathbf{\Gamma}_z = \widetilde{\mathbf{F}}, \tag{A.9}$$
$$\mathbf{\Upsilon}_x = \widetilde{\mathbf{Q}}, \quad \mathbf{\Upsilon}_z = \widetilde{\mathbf{F}}, \tag{A.10}$$

and we obtain the finite-state representation. Note that the eigenvalues of $\mathbf{\Gamma}_z$ all lie inside the unit circle. The law of motion of $\mathbf{z}_t$ can be written as

$$\mathbf{z}_t = (\mathbf{I} - \mathbf{\Upsilon}_z L)^{-1} \mathbf{\Upsilon}_x \mathbf{x}_t \tag{A.11}$$

Therefore, given $\{\mathbf{x}_t\}_{t=-\infty}^{-1}$,

$$\mathbf{z}_{-1} = (\mathbf{I} - \mathbf{\Upsilon}_z L)^{-1} \mathbf{\Upsilon}_x \mathbf{x}_{-1} \tag{A.12}$$

□

A.9 Proof of Theorem 3.4

Proof. Suppose there exists a solution in signal form

$$\mathbf{y}_t = \mathbf{h}(L)\mathbf{x}_t.$$

By the definition of the signal process (3.1), it follows that

$$\mathbf{y}_t = \mathbf{h}(L)\mathbf{M}(L)\mathbf{s}_t.$$

Because $\mathbf{y}_t = \mathbf{h}(L)\mathbf{x}_t$ is a solution to model (3.4), $\mathbf{y}_t = \mathbf{h}(L)\mathbf{M}(L)\mathbf{s}_t$ represents the same process and is also a solution model (3.4).

Reversely, suppose there exists a solution in innovation form

$$\mathbf{y}_t = \mathbf{d}(L)\mathbf{s}_t.$$

Since $\mathbf{y}_t$ always lies in the space spanned by current and past signals, it follows that

$$\mathbf{d}(L)\mathbf{s}_t = \mathbb{E}_t[\mathbf{d}(L)\mathbf{s}_t] = \mathbf{h}(L)\mathbf{x}_t.$$

By Proposition 3.1,

$$\mathbf{vec}(\mathbf{h}(L)) = \mathbf{vec}(\boldsymbol{\rho}'_{xx}(L)^{-1}\mathbf{M}(L^{-1})\mathbf{d}'(L)) - \sum_{k=1}^{u} \frac{\mathbf{vec}(\mathbf{B}'(L)^{-1}\mathbf{V}^{-1}\mathbf{G}'(\lambda_k)\mathbf{d}'(\lambda_k))}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k - \lambda_\tau)}.$$

□

A.10 Proof of Proposition 4.1

Proof. By the Canonical Factorization Theorem, the prior variance of the state ξ_t and the Kalman gain matrix satisfies

$$\begin{aligned}
\mathbf{P} &= \mathbf{F}[\mathbf{P} - \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1}\mathbf{H}\mathbf{P}]\mathbf{F}' + \mathbf{\Phi}\mathbf{\Phi}' \\
\mathbf{K} &= \mathbf{P}\mathbf{H}'(\mathbf{H}\mathbf{P}\mathbf{H}' + \mathbf{\Psi}\mathbf{\Psi}')^{-1}
\end{aligned}$$

Using the state-space representation in equation (4.5), $\mathbf{P}$ will be a scalar. Denote $\kappa \equiv \mathbf{P}^{-1}$ as the prior precision about ξ_t, it is easy to verify that

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

The forecast about the fundamental ξ_t is given by

$$\mathbb{E}_{it}[\xi_t] = \sigma_u^2\sigma_\epsilon^2\kappa(\sigma_u^2\sigma_\epsilon^2\kappa + \sigma_u^2 + \sigma_\epsilon^2)^{-1}\rho\mathbb{E}_{it-1}[\xi_{t-1}] + (\sigma_u^2\sigma_\epsilon^2\kappa + \sigma_u^2 + \sigma_\epsilon^2)^{-1}(\sigma_u^2 x_{it}^1 + \sigma_\epsilon^2 x_{it}^2)$$

Define $\tau_1 = \frac{\sigma_\epsilon^2}{\sigma_\eta^2}$, $\tau_1 = \frac{\sigma_u^2}{\sigma_\eta^2}$, and $\lambda = \sigma_u^2\sigma_\epsilon^2\kappa(\sigma_u^2\sigma_\epsilon^2\kappa + \sigma_u^2 + \sigma_\epsilon^2)^{-1}\rho$, and it follows that

$$\lambda = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{\tau_1+\tau_2}{\rho\tau_1\tau_2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\tau_1+\tau_2}{\rho\tau_1\tau_2}\right)^2 - 4}\right]$$

The Wold representation is

$$\begin{aligned}
\mathbf{B}(z)^{-1} &= \frac{1}{1-\lambda z}\begin{bmatrix} 1 - \frac{\tau_2\rho+\lambda\tau_1}{\tau_1+\tau_2}z & \frac{\tau_1(\lambda-\rho)}{\tau_1+\tau_2}z \\ \frac{\tau_2(\lambda-\rho)}{\tau_1+\tau_2}z & 1 - \frac{\tau_1\rho+\lambda\tau_2}{\tau_1+\tau_2}z \end{bmatrix}, \\
\mathbf{V}^{-1} &= \frac{1}{\rho(\tau_1+\tau_2)}\begin{bmatrix} \frac{\tau_1\rho+\lambda\tau_2}{\tau_1} & \lambda-\rho \\ \lambda-\rho & \frac{\tau_2\rho+\lambda\tau_1}{\tau_2} \end{bmatrix},
\end{aligned}$$

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

$$y_t = (h_1(L) + h_2(L))\xi_t + h_1(L)\epsilon_t.$$

By Proposition 3.1, we have

$$\mathbb{E}_{it}[\xi_t] = \begin{bmatrix} \frac{1}{1-\lambda L}\frac{\lambda}{(1-\rho\lambda)\rho\tau_1} \\ \frac{1}{1-\lambda L}\frac{\lambda}{(1-\rho\lambda)\rho\tau_2} \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix},$$

and

$$\mathbb{E}_{it}[y_t] = \begin{bmatrix} \frac{\lambda}{\rho\tau_1}\frac{L}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \\ \frac{\lambda}{\rho\tau_2}\frac{L}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2}{\rho\tau_2}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{\lambda}{\rho\tau_1}\frac{L}{(1-\lambda L)(L-\lambda)}h_2(L) - \frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \\ \frac{\lambda}{\rho\tau_2}\frac{L}{(1-\lambda L)(L-\lambda)}h_2(L) - \frac{\lambda^2}{\rho\tau_2}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{\tau_1}{\tau_1+\tau_2}h_1(L) + \frac{\tau_2\frac{\lambda}{\rho}(L-\rho)(1-\rho L)}{(\tau_1+\tau_2)(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\tau_2\frac{\lambda}{\rho}(\lambda-\rho)(1-\rho L)}{(\tau_1+\tau_2)(1-\lambda L)(L-\lambda)}h_1(\lambda) \\ -\frac{\tau_1}{\tau_1+\tau_2}h_1(L) + \frac{\tau_1\frac{\lambda}{\rho}(L-\rho)(1-\rho L)}{(\tau_1+\tau_2)(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\tau_1\frac{\lambda}{\rho}(\lambda-\rho)(1-\rho L)}{(\tau_1+\tau_2)(1-\lambda L)(L-\lambda)}h_1(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix}$$

The model requires that

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha\mathbb{E}_{it}[y_t],$$

which leads to the following system of analytic functions

$$C(z)\begin{bmatrix} h_1(z) \\ h_2(z) \end{bmatrix} = d(z, h(\lambda))$$

where $h(\lambda) = h_2(\lambda)$, and

$$C(z) = \begin{bmatrix} 1 - \alpha\frac{\lambda}{\rho\tau_1}\frac{z}{(1-\lambda z)(z-\lambda)} - \alpha\left(\frac{\tau_1}{\tau_1+\tau_2} + \frac{\tau_2\frac{\lambda}{\rho}(z-\rho)(1-\rho z)}{(\tau_1+\tau_2)(1-\lambda z)(z-\lambda)}\right) & -\alpha\frac{\lambda}{\rho\tau_1}\frac{z}{(1-\lambda z)(z-\lambda)} \\ -\alpha\frac{\lambda}{\rho\tau_2}\frac{z}{(1-\lambda z)(z-\lambda)} - \alpha\left(-\frac{\tau_1}{\tau_1+\tau_2} + \frac{\tau_1\frac{\lambda}{\rho}(z-\rho)(1-\rho z)}{(\tau_1+\tau_2)(1-\lambda z)(z-\lambda)}\right) & 1 - \alpha\frac{\lambda}{\rho\tau_2}\frac{z}{(1-\lambda z)(z-\lambda)} \end{bmatrix},$$

$$d(z, h(\lambda)) = \begin{bmatrix} d_1(z, h(\lambda)) \\ d_2(z, h(\lambda)) \end{bmatrix} = \begin{bmatrix} \frac{1}{1-\lambda z}\frac{\lambda}{(1-\rho\lambda)\rho\tau_1} - \alpha\frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}\frac{1-\rho z}{(1-\lambda z)(z-\lambda)}h(\lambda) \\ \frac{1}{1-\lambda z}\frac{\lambda}{(1-\rho\lambda)\rho\tau_2} - \alpha\frac{\lambda^2}{\rho\tau_2}\frac{1}{1-\rho\lambda}\frac{1-\rho z}{(1-\lambda z)(z-\lambda)}h(\lambda) \end{bmatrix}.$$

Note that

$$\det C(z) = \frac{(1-\alpha)\lambda(z-\vartheta)(1-\vartheta z)}{\vartheta(1-\lambda z)(z-\lambda)}$$

The inside root of the determinant of $C(z)$ is

$$\vartheta = \frac{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\tau_1+\tau_2}{\rho\tau_1\tau_2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\tau_1+\tau_2}{\rho\tau_1\tau_2}\right)^2 - 4}}{2}$$

Using Cramer's rule,

$$h_1(z) \quad = \quad \frac{\det \begin{bmatrix} d_1(z) & -\alpha \frac{\lambda}{\rho\tau_1} \frac{z}{(1-\lambda z)(z-\lambda)} \\ \\ d_2(z) & 1-\alpha \frac{\lambda}{\rho\tau_2} \frac{z}{(1-\lambda z)(z-\lambda)} \end{bmatrix}}{\det C(z)}.$$

The numerator is

$$\begin{aligned} &\det \begin{bmatrix} d_1(z) & -\alpha \frac{\lambda}{\rho\tau_1} \frac{z}{(1-\lambda z)(z-\lambda)} \\ \\ d_2(z) & 1-\alpha \frac{\lambda}{\rho\tau_2} \frac{z}{(1-\lambda z)(z-\lambda)} \end{bmatrix} \\ = \quad & \frac{1}{(1-\lambda z)(z-\lambda)} \left\{ \frac{\lambda(z-\lambda)}{(1-\rho\lambda)\rho\tau_1} - \alpha \frac{\lambda^2}{\rho\tau_1} \frac{1}{1-\rho\lambda}(1-\rho z)h(\lambda) \right\}. \end{aligned}$$

To make sure $h_1(z)$ does not have poles in the unit circle, we need to choose $h(\lambda)$ to remove the pole at ϑ, which requires

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

Therefore,

$$h_1(z) = \frac{\vartheta}{\rho\tau_1(1-\alpha)(1-\rho\vartheta)} \frac{1}{1-\vartheta z},$$

and similarly,

$$h_2(z) = \frac{\vartheta}{\rho\tau_1(1-\rho\vartheta)} \frac{1}{1-\vartheta z}$$

$\square$

A.11 Proof of Proposition 4.3

Proof. The signal process is

$$x_{it} = \mathbf{M}(L)\mathbf{s}_{it} = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & \sigma_u \end{bmatrix} \begin{bmatrix} \eta_t \\ u_{it} \end{bmatrix}$$

The corresponding fundamental representation is

$$B(z) = \frac{1-\lambda z}{1-\lambda z}, \quad V = \frac{\sigma_u^2 \rho}{\lambda}$$

where λ is

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

which satisfy

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

It is straightforward to verify that

$$\mathbf{G}(z) = \begin{bmatrix} \sigma_\eta z \\ \sigma_u \end{bmatrix}$$

Forward-looking Model We start with the forward-looking model in which

$$P(L) = 1, \quad Q(L) = -\alpha L^{-1}, \quad \mathbf{R}(L) = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & 0 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

The system of equation is

$$\begin{aligned} T(z) &= \frac{-\rho\sigma_u^2\left(z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right)z + 1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right)}{(1-\rho z)(z-\rho)} = \frac{-\rho\sigma_u^2(z-\zeta_1)(z-\zeta_2)}{(1-\rho z)(z-\rho)} \\ D(z, h(\lambda)) &= \frac{\sigma_\eta^2(z-\lambda)}{(1-\rho\lambda)(1-\rho z)(z-\rho)} - \alpha\frac{\sigma_\eta^2\lambda}{(1-\rho\lambda)(z-\rho)}h(\lambda) \end{aligned}$$

To have a unique solution, it has to be that $|\zeta_1| < 1$ and $|\zeta_2| > 1$. Note that when $\alpha < 1$,

$$z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right)z + 1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\bigg|_{z=1} = 2 - \left(\rho + \frac{1}{\rho}\right) - (1-\alpha)\frac{\sigma_\eta^2}{\rho\sigma_u^2} < 0$$

which guarantees there exists a unique solution. Multiply $(1-\rho z)(z-\rho)$ to both sides and set $h(\lambda)$ such that

$$\frac{\sigma_\eta^2(z-\lambda)}{1-\rho\lambda} - \alpha\frac{\sigma_\eta^2\lambda(1-\rho z)}{(1-\rho\lambda)}h(\lambda) = 0\bigg|_{z=\zeta_1}$$

This leads to that

$$h(z) = -\frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho(1-\rho\zeta_1)}\frac{1}{z-\zeta_2}$$

Defining $\vartheta_f \equiv \zeta_2^{-1}$, it follows that

$$h(z) = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_f}{\rho\left(1-\rho\vartheta_f - \alpha\frac{\sigma_\eta^2}{\sigma_u^2}\vartheta_f\right)}\frac{1}{1-\vartheta_f z} \tag{A.13}$$

where

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

Static Model The static model is the same as Application I. The result is that

$$h(z) = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_s}{\rho(1-\rho\vartheta_s)}\frac{1}{1-\vartheta_s z}$$

where

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

Backward-looking Model In the backward-looking model,

$$P(L) = 1, \quad Q(L) = -\alpha L, \quad \mathbf{R}(L) = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & 0 \end{bmatrix}, \quad \mathbf{\Lambda} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

The system of equation is

$$T(z) = \frac{-(\rho\sigma_u^2+\alpha\sigma_\eta^2)\left(z^2 - \frac{\sigma_\eta^2+\sigma_u^2(1+\rho^2)}{\rho\sigma_u^2+\alpha\sigma_\eta^2}z + \frac{\rho\sigma_u^2}{\rho\sigma_u^2+\alpha\sigma_\eta^2}\right)}{(1-\rho z)(z-\rho)} = \frac{-(\rho\sigma_u^2+\alpha\sigma_\eta^2)(z-\zeta_1)(z-\zeta_2)}{(1-\rho z)(z-\rho)}$$

$$D(z, h(\lambda)) = \frac{\sigma_\eta^2(z-\lambda)}{(1-\rho\lambda)(1-\rho z)(z-\rho)} - \alpha\frac{\sigma_\eta^2\lambda^2}{(1-\rho\lambda)(z-\rho)}h(\lambda)$$

To have a unique solution, it has to be that $|\zeta_1| < 1$ and $|\zeta_2| > 1$. Note that when $\alpha \in (0,1)$,

$$z^2 - \frac{\sigma_\eta^2+\sigma_u^2(1+\rho^2)}{\rho\sigma_u^2+\alpha\sigma_\eta^2}z + \frac{\rho\sigma_u^2}{\rho\sigma_u^2+\alpha\sigma_\eta^2}\bigg|_{z=1} = -\frac{(1-\alpha)\sigma_\eta^2 + \rho\sigma_u^2\left(\rho+\frac{1}{\rho}-2\right)}{\rho\sigma_u^2+\alpha\sigma_\eta^2} < 0$$

which guarantees there exists a unique solution. Multiply $(1-\rho z)(z-\rho)$ to both sides and set $h(\lambda)$ such that

$$\frac{\sigma_\eta^2(z-\lambda)}{1-\rho\lambda} - \alpha\frac{\sigma_\eta^2\lambda^2(1-\rho z)}{(1-\rho\lambda)}h(\lambda) = 0\bigg|_{z=\zeta_1}$$

This leads to that

$$h(z) = -\frac{\sigma_\eta^2}{(\rho\sigma_u^2+\alpha\sigma_\eta^2)}\frac{1}{1-\rho\zeta_1}\frac{1}{z-\zeta_2}$$

Defining $\vartheta_b \equiv \zeta_2^{-1}$, it follows that

$$h(z) = \frac{\vartheta_b\sigma_\eta^2}{\rho\sigma_u^2+\alpha\sigma_\eta^2-\rho^2\sigma_u^2\vartheta_b}\frac{1}{1-\vartheta_b z} \tag{A.14}$$

where

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

Comparison Recall that ϑ_f^{-1} is a root of

$$z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right) z + 1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}$$

while ϑ_s^{-1} is a root of

$$z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho - \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right) z + 1$$

Note that

$$z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right) z + 1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\bigg|_{z=\vartheta_s^{-1}} = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}(1-\vartheta_s^{-1}) < 0$$

Therefore, $\vartheta_f < \vartheta_s$.

To compare ϑ_s and ϑ_b, recall that ϑ_b^{-1} is a root of

$$z^2 - \frac{\sigma_\eta^2 + \sigma_u^2(1+\rho^2)}{\rho\sigma_u^2 + \alpha\sigma_\eta^2} z + \frac{\rho\sigma_u^2}{\rho\sigma_u^2 + \alpha\sigma_\eta^2} = \frac{\rho\sigma_u^2}{\rho\sigma_u^2 + \alpha\sigma_\eta^2}\left(\left(1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right) z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right) z + 1\right)$$

Note that

$$\left(1 + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\right) z^2 - \left(\frac{1}{\rho} + \frac{\sigma_\eta^2}{\sigma_u^2}\frac{1}{\rho} + \rho\right) z + 1\bigg|_{z=\vartheta_s^{-1}} = \frac{\sigma_\eta^2}{\sigma_u^2}\frac{\alpha}{\rho}\vartheta_s^{-1}(\vartheta_s^{-1} - 1) > 0$$

Therefore, $\vartheta_s < \vartheta_b$. □

A.12 Proof of Proposition 4.5

Proof. For agent i, the signal process is

$$x_{it} = \mathbf{M}(L)\mathbf{s}_{it} = \begin{bmatrix} \frac{\sigma_\eta}{1-\rho L} & \sigma_u \end{bmatrix} \begin{bmatrix} \eta_t \\ u_{it} \end{bmatrix}$$

Similar to the Proof A.11, the corresponding fundamental representation is

$$B(z) = \frac{1-\lambda z}{1-\lambda z}, \quad V = \frac{\sigma_u^2 \rho}{\lambda}$$

where λ is

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

which satisfy

$$\mathbf{M}(z)\mathbf{M}'(z^{-1}) = B(z)VB'(z^{-1})$$

It is straightforward to verify that

$$\mathbf{G}(z) = \begin{bmatrix} \sigma_\eta z \\ \sigma_u \end{bmatrix}$$

Now consider agent i's forecast problem. Let j be the index of the other agent, and the forecasts are given by

$$\begin{aligned}
\mathbb{E}_{it}[\xi_t] &= \frac{\sigma_\eta^2}{(1-\rho\lambda)} V^{-1} \frac{1}{1-\lambda L} x_{it} \\
\mathbb{E}_{it}[y_{jt}] &= \left(\frac{\sigma_\eta^2 h_j(L) L}{L-\lambda} - \frac{\sigma_\eta^2 h_j(\lambda)\lambda(1-\rho L)}{(1-\rho\lambda)(L-\lambda)} \right) V^{-1} \frac{1}{1-\lambda L} x_{it}
\end{aligned}$$

The best response requires that

$$V(z-\lambda)(1-\lambda z) h_i(z) - \omega_i \sigma_\eta^2 z h_j(z) = (1-\omega_i) \frac{\sigma_\eta^2 (z-\lambda)}{(1-\rho\lambda)} - \omega_i \frac{\sigma_\eta^2 h_j(\lambda)\lambda(1-\rho z)}{(1-\rho\lambda)}$$

For agent j, a similar equilibrium equation can be derived. Combining the two agents' best response leads to

$$\mathbf{T}(z) \begin{bmatrix} h_1(z) \\ h_2(z) \end{bmatrix} = \mathbf{D}(z, h_1(\lambda_2), h_2(\lambda_1))$$

where

$$\mathbf{T}(z) = \begin{bmatrix} V(z-\lambda)(1-\lambda z) & -\omega_1 \sigma_\eta^2 z \\ -\omega_2 \sigma_\eta^2 z & V(z-\lambda)(1-\lambda z) \end{bmatrix}$$

$$\mathbf{D}(z, h_1(\lambda), h_2(\lambda)) = \begin{bmatrix} \frac{\sigma_\eta^2 (z-\lambda)}{(1-\rho\lambda)} - \omega_1 \frac{\sigma_\eta^2 h_2(\lambda)\lambda(1-\rho z)}{(1-\rho\lambda)} \\ \frac{\sigma_\eta^2 (z-\lambda)}{(1-\rho\lambda)} - \omega_2 \frac{\sigma_\eta^2 h_1(\lambda)\lambda(1-\rho z)}{(1-\rho\lambda)} \end{bmatrix}$$

The determinant of $\mathbf{T}(z)$ is given by

$$\begin{aligned}
\det[\mathbf{T}(z)] &= V^2 (z-\lambda)(1-\lambda z)(z-\lambda)(1-\lambda z) - \omega_1\omega_2 \sigma_\eta^4 z^2 \\
&= \sigma_u^4 \rho^2 (z-\vartheta_1)(z-\vartheta_1^{-1})(z-\vartheta_2)(z-\vartheta_2^{-1}) \\
&= \frac{\sigma_u^4 \rho^2}{\vartheta_1 \vartheta_2} (z-\vartheta_1)(1-\vartheta_1 z)(z-\vartheta_2)(1-\vartheta_2 z)
\end{aligned}$$

We can characterize ϑ_1 and ϑ_2 in the following way. First notice that

$$\left(\vartheta_1 + \frac{1}{\vartheta_1}\right) + \left(\vartheta_2 + \frac{1}{\vartheta_2}\right) = 2\left(\lambda + \frac{1}{\lambda}\right)$$

$$\left(\vartheta_1+\frac{1}{\vartheta_1}\right)\left(\vartheta_2+\frac{1}{\vartheta_2}\right)=\left(\lambda+\frac{1}{\lambda}\right)^2-\frac{\omega_1\omega_2\sigma_\eta^4}{\rho^2\sigma_u^4}$$

then we have

$$\vartheta_1+\frac{1}{\vartheta_1}=\left(\lambda+\frac{1}{\lambda}\right)-\frac{\sqrt{\omega_1\omega_2}\sigma_\eta^2}{\rho\sigma_u^2},\quad \vartheta_2+\frac{1}{\vartheta_2}=\left(\lambda+\frac{1}{\lambda}\right)+\frac{\sqrt{\omega_1\omega_2}\sigma_\eta^2}{\rho\sigma_u^2}$$

and

$$\vartheta_1=\frac{1}{2}\left[\left(\frac{1}{\rho}+\rho+\frac{(1-\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)-\sqrt{\left(\frac{1}{\rho}+\rho+\frac{(1-\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)^2-4}\right]$$
$$\vartheta_2=\frac{1}{2}\left[\left(\frac{1}{\rho}+\rho+\frac{(1+\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)-\sqrt{\left(\frac{1}{\rho}+\rho+\frac{(1+\sqrt{\omega_1\omega_2})\sigma_\eta^2}{\rho\sigma_u^2}\right)^2-4}\right]$$

To determine $h_1(\lambda)$ and $h_2(\lambda)$, we need to set

$$\det\begin{bmatrix}\frac{\sigma_\eta^2(z-\lambda)}{(1-\rho\lambda)}-\omega_1\frac{\sigma_\eta^2h_2(\lambda)\lambda(1-\rho z)}{(1-\rho\lambda)} & -\omega_1\sigma_\eta^2 z\\ \frac{\sigma_\eta^2(z-\lambda)}{(1-\rho\lambda)}-\omega_2\frac{\sigma_\eta^2h_1(\lambda)\lambda(1-\rho z)}{(1-\rho\lambda)} & V(z-\lambda)(1-\lambda z)\end{bmatrix}=0$$

when evaluated at $z=\vartheta_1$ and $z=\vartheta_2$. After some algebra, we have

$$h_1(L)=\frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_1\vartheta_2}{\rho(1-\vartheta_1\vartheta_2)}\frac{\frac{1}{\rho}\left((1-\omega_1)+\frac{(1-\omega_1\omega_2)}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)-\left((1-\omega_1)+\frac{(1-\omega_1\omega_2)\vartheta_1\vartheta_2}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)L}{(1-\vartheta_1L)(1-\vartheta_2L)}$$
$$h_2(L)=\frac{\sigma_\eta^2}{\sigma_u^2}\frac{\vartheta_1\vartheta_2}{\rho(1-\vartheta_1\vartheta_2)}\frac{\frac{1}{\rho}\left((1-\omega_2)+\frac{(1-\omega_1\omega_2)}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)-\left((1-\omega_2)+\frac{(1-\omega_1\omega_2)\vartheta_1\vartheta_2}{(1-\rho\vartheta_1)(1-\rho\vartheta_2)}\frac{\sigma_\eta^2}{\sigma_u^2}\right)L}{(1-\vartheta_1L)(1-\vartheta_2L)}$$

□

A.13 Proof of Proposition 4.4

Proof. Note that $x^1_{m(i,t)t}=a_i+\epsilon_{m(i,t)t}$, the signal process can be equivalently rewritten as

$$\begin{aligned}x^1_{it}&=a_{m(i,t)}+\sigma_\epsilon\epsilon_{it}\\ \widehat{x}^2_{it}&=x^2_{m(i,t)t}-a_i=\xi_t+\sigma_\epsilon\epsilon_{m(i,t)t}+\sigma_u u_{it},\\ \xi_t&=\rho\xi_{t-1}+\eta_t.\end{aligned}$$

Denote the policy rule using this transformed signals as

$$y_{it}=g_a a_i+g_1(L)x^1_{it}+g_2(L)\widehat{x}^2_{it}$$

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

$$
\begin{aligned}
h_a &= g_a - g_2(1) \\
h_1(L) &= g_1(L) \\
h_2(L) &= g_2(L).
\end{aligned}
$$

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

$$
\begin{aligned}
B(z) &= \frac{1-\lambda z}{1-\rho z}, \\
V &= \frac{\rho(\sigma_\epsilon^2+\sigma_u^2)}{\lambda\sigma_\eta^2},
\end{aligned}
$$

where

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

The prediction of $y_{m(i,t)t}$ is

$$
\mathbb{E}_{it}[y_{m(i,t)t}] = \mathbb{E}_{it}[g_a a_{m(i,t)} + g_1(L)(a_{m(m(i,t),t)} + \sigma_\epsilon \epsilon_{m(i,t)t}) + g_2(L)(\sigma_u u_{(m(i,t)t)} + \sigma_\epsilon \epsilon_{m(m(i,t),t)} + \xi_t)],
$$

where

$$
\begin{aligned}
\mathbb{E}_{it}[a_{m(i,t)}] &= \frac{\sigma_a^2}{\sigma_a^2+\sigma_\epsilon^2} x^1_{it} \\
\mathbb{E}_{it}[a_{m(m(i,\tau),\tau)}] &= a_i \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[\sigma_\epsilon \epsilon_{m(i,\tau)\tau}] &= \frac{\lambda\sigma_\epsilon^2\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)} \frac{1-\rho L}{1-\lambda L} \widehat{x}^2_{it} \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[u_{m(i,t)t}] &= 0 \\
\mathbb{E}_{it}[\sigma_\epsilon \epsilon_{m(m(i,\tau),\tau)}] &= \frac{\sigma_\epsilon^2}{\sigma_a^2+\sigma_\epsilon^2} x^1_{it} \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[g_2(L)\xi_t] &= \left(\frac{Lg_2(L)}{L-\lambda} - \frac{\lambda(1-\rho L)g_2(\lambda))}{(1-\rho\lambda)(L-\lambda)}\right) \frac{V^{-1}}{1-\lambda L} \widehat{x}^2_{it}.
\end{aligned}
$$

The best response requires that

$$
\begin{aligned}
& g_a a_i + g_1(L) x^1_{it} + g_2(L) \widehat{x}^2_{it} \\
=& a_i + \alpha \Bigg[g_a \frac{\sigma_a^2}{\sigma_a^2+\sigma_\epsilon^2} x^1_{it} + g_1(0) a_i + g_1(0) \frac{\lambda\sigma_\epsilon^2\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)} \frac{1-\rho L}{1-\lambda L} \widehat{x}^2_{it}
\end{aligned}
$$

$$+\left(\frac{Lg_2(L)}{L-\lambda}-\frac{\lambda(1-\rho L)g_2(\lambda))}{(1-\rho\lambda)(L-\lambda)}\right)\frac{V^{-1}}{1-\lambda L}\widehat{x}_{it}^2+g_2(0)\frac{\sigma_\epsilon^2}{\sigma_a^2+\sigma_\epsilon^2}x_{it}^1\Big],$$

which leads to

$$\begin{aligned}
g_a &= 1+\alpha g_1(0)\\
g_1(0) &= \alpha g_a\frac{\sigma_a^2}{\sigma_a^2+\sigma_\epsilon^2}+\alpha g_2(0)\frac{\sigma_\epsilon^2}{\sigma_a^2+\sigma_\epsilon^2}\\
g_2(z) &= \alpha g_1(0)\frac{\lambda\sigma_\epsilon^2\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}\frac{1-\rho z}{1-\lambda z}+\alpha\left(\frac{zh_2(z)}{z-\lambda}-\frac{\lambda(1-\rho z)h_2(\lambda))}{(1-\rho\lambda)(z-\lambda)}\right)\frac{V^{-1}}{1-\lambda z}.
\end{aligned}$$

The third equation can be written as

$$-(z-\vartheta)\left(z-\frac{1}{\vartheta}\right)g_2(z)=\alpha g_1(0)\frac{\sigma_\epsilon^2\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}(1-\rho z)(z-\lambda)-\alpha\frac{V^{-1}(1-\rho z)g_2(\lambda)}{(1-\rho\lambda)}$$

where

$$\vartheta=\frac{1}{2}\left[\frac{1}{\rho}+\rho+\frac{(1-\alpha)\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}-\sqrt{\left(\frac{1}{\rho}+\rho+\frac{(1-\alpha)\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}\right)^2-4}\right]. \tag{A.15}$$

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

$$\begin{aligned}
g_1(z) = g_1(0) &= \frac{\alpha}{1-\alpha^2+\frac{\sigma_\epsilon^2}{\sigma_a^2}\left(1-\alpha^2\frac{\vartheta}{\rho}\frac{\sigma_\epsilon^2\sigma_\eta^2}{\sigma_\epsilon^2+\sigma_u^2}\right)}\\
g_a &= 1+\alpha g_1(0)\\
g_2(z) &= \frac{\alpha\vartheta g_1(0)\sigma_\epsilon^2\sigma_\eta^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}\frac{1-\rho z}{1-\vartheta z}
\end{aligned}$$

□

A.14 Proof of Proposition 5.1

Proof. Let $\phi=\{\phi_1,\phi_2,\phi_3\}\in\ell^2\times\ell^2\times\ell^2$ denote an arbitrary policy rule. The norm of ϕ can de defined as

$$\|\phi\|=\sqrt{\sigma_\epsilon^2\sum_{k=0}^\infty\phi_{1k}^2+\sigma_u^2\sum_{k=0}^\infty\phi_{2k}^2+\sigma_\eta^2\sum_{k=0}^\infty\phi_{3k}^2}.$$

The individual action is then given by

$$y_{it}=\phi_\epsilon(L)\epsilon_{it}+\phi_u(L)u_{it}+\phi_\eta(L)\eta_t.$$

Given ϕ, define the signal process as

$$\begin{aligned} x_{it}^1 &= \xi_t + \sigma_\epsilon \epsilon_{it}, \\ x_{it}^2 &= \phi_\eta(L)\eta_t + \sigma_u u_{it}, \end{aligned}$$

where we have already used the equilibrium condition that $y_t = \phi_\eta(L)\eta_t$. The optimal linear forecast conditional on the signals is denoted as

$$\mathbb{E}_{it}[y_t] = \widehat{\phi}_\epsilon(L)\epsilon_{it} + \widehat{\phi}_u(L)u_{it} + \widehat{\phi}_\eta(L)\eta_t.$$

Note that the individual optimality requires that

$$\phi_\epsilon(L)\epsilon_{it} + \phi_u(L)u_{it} + \phi_\eta(L)\eta_t = \xi_t + \sigma_\epsilon \epsilon_{it} + \alpha\mathbb{E}_{it}[y_t].$$

Define the operator $\mathcal{T} : \ell^2 \times \ell^2 \times \ell^2 \to \ell^2 \times \ell^2 \times \ell^2$ as

$$\mathcal{T}(\phi) = \{\sigma_\epsilon + \alpha\widehat{\phi}_\epsilon, \alpha\widehat{\phi}_u, \sigma_\eta\phi_\xi + \alpha\widehat{\phi}_\eta\}$$

where $\phi_\xi \equiv \{1, \rho, \rho^2, \ldots\}$ denotes the coefficients of the lag polynomial of ξ_t. The equilibrium is a fixed point of the operator $\mathcal{T}$. For the existence of the fixed point, we will rely on the Schauder's fixed point theorem. To apply this theorem, we need to show that ϕ always belongs to a compact space. It turns out a higher order belief representation of individual's action is sufficient to prove it. By consecutive iteration, we have

$$\begin{aligned} y_{it} &= \xi_{it} + \alpha\mathbb{E}_{it}[y_t] \\ &= \xi_{it} + \alpha\mathbb{E}_{it}\left[\int y_{jt}\right] \\ &= \xi_{it} + \alpha\mathbb{E}_{it}\left[\xi_t\right] + \alpha^2\mathbb{E}_{it}\left[\int \mathbb{E}_{jt}[y_{jt}]\right] \\ &\vdots \\ &= \xi_{it} + \sum_{k=1}^{\infty} \alpha^k \mathbb{E}_{it}^k[\xi_t] \end{aligned}$$

By the law of total variance, the variance of $\mathbb{E}_{it}^k[\xi_t]$ is less than the variance of ξ_t. Therefore, no matter what the signal process is, the variance of y_{it} is bounded. Therefore, the policy rule has to belong to a compact space bounded by the norm of ξ_t, and the operator $\mathcal{T}$ is a bounded continuous operator. This completes the proof.

□

A.15 Proof of Theorem 2

Proof. We first layout the structure of the proof, then we enter the details of each step.

1. Assume the law of aggregate y_t has a finite ARMA representation in condition 1 of definition 5.1.

$$\varphi(L) = \frac{a(L)}{b(L)}, \tag{A.16}$$

   where $a(L)$ and $b(L)$ are finite polynomials in L.

2. Solve agents optimal policy $\boldsymbol{\phi} = \{\phi_1, \phi_2, \phi_3\}$ in a partial equilibrium with exogenous information. The partial equilibrium consists of two conditions

   - Each individual makes inference conditional on the following signal process

$$\begin{aligned} x_{it}^1 &= \xi_t + \sigma_\epsilon \epsilon_{it} \\ x_{it}^2 &= \varphi(L)\eta_t + \sigma_u u_{it} \end{aligned}$$

   - The policy rule $\boldsymbol{\phi}$ satisfies that

$$y_{it} = \xi_{it} + \alpha\, \mathbb{E}_{it}\left[y_t\right],$$

     where

$$\begin{aligned} y_{it} &= \phi_\epsilon(L)\epsilon_{it} + \phi_u(L)u_{it} + \phi_\eta(L)\eta_t, \\ y_t &= \phi_\eta(L)\eta_t. \end{aligned}$$

   Note that in this partial equilibrium, agents reply on exogenous information, but their optimal policy rule does depend on others' action. Also note that we do not require $\phi_\eta(L) = \varphi(L)$. Solving this partial equilibrium is similar to the problem in Section 3.

3. Show $\varphi(L)$ cannot be the same as $\phi_\eta(L)$. That is, condition 3 of definition 5.1 cannot be satisfied.

Now we move to the details of each step.

Step 1 Under the finite ARMA representation assumption, the signal process is given by

$$\mathbf{x}_{it} = \mathbf{M}(L)\mathbf{s}_{it} = \begin{bmatrix} \sigma_\epsilon & 0 & \frac{1}{1-\rho L} \\ 0 & \sigma_u & \frac{a(L)}{b(L)} \end{bmatrix} \begin{bmatrix} \epsilon_{it} \\ u_{it} \\ \eta_t \end{bmatrix}$$

The state-space and fundamental representation are

$$
\begin{aligned}
\mathbf{x}_{it} &= \mathbf{H}(\mathbf{I}-\mathbf{F}L)^{-1}\mathbf{Q} \\
\mathbf{M}(L)\mathbf{M}'(L^{-1}) &= \mathbf{B}(L)\mathbf{V}^{-1}\mathbf{B}'(L^{-1})
\end{aligned}
$$

Step 2 By Proposition 3.1, the forecast about $y_t = \phi_\eta(L)\eta_t$ is given by

$$\mathbb{E}_{it}[\phi_\eta(L)\eta_t] = \left[\begin{bmatrix}0 & 0 & \phi_\eta(L)\end{bmatrix}\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_+ \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{M}(L)\mathbf{s}_{it}.$$

As shown in the proof A.4, there exists finite degree polynomial matrices $\mathbf{G}(L)$ and $\mathbf{K}(L)$ such that

$$\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1} = \frac{\mathbf{G}(L)}{\prod_{k=1}^{d}(L-\lambda_i)} \tag{A.17}$$

$$\mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{M}(L) = \frac{\mathbf{K}(L)}{\prod_{k=1}^{d}(1-\lambda_i L)} \tag{A.18}$$

where $\{\lambda_i\}_{i=1}^{d}$ are non-zero eigenvalues of $\mathbf{F}-\mathbf{FKH}$ in the associated steady-state Kalman filter problem. The forecast about the aggregate action in the partial equilibrium is

$$
\begin{aligned}
\mathbb{E}_{it}[y_t] = \mathbb{E}_{it}[\phi_\eta(L)\eta_t] &= \left[\begin{bmatrix}0 & 0 & \phi_\eta(L)\end{bmatrix}\mathbf{M}'(L^{-1})\mathbf{B}'(L^{-1})^{-1}\right]_+ \mathbf{V}^{-1}\mathbf{B}(L)^{-1}\mathbf{M}(L)\mathbf{s}_{it} \\
&= \phi_\eta(L)\begin{bmatrix}0 & 0 & 1\end{bmatrix}\mathbf{M}'(L^{-1})\boldsymbol{\rho}_{xx}(L)^{-1}\mathbf{M}(L)\mathbf{s}_{it} - \sum_{k=1}^{d}\frac{\begin{bmatrix}0 & 0 & \phi_\eta(\lambda_k)\end{bmatrix}\mathbf{G}(\lambda_k)}{(L-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)}\frac{\mathbf{K}(L)}{\prod_{k=1}^{d}(1-\lambda_i L)}\mathbf{s}_{it}
\end{aligned}
$$

The partial equilibrium condition for $\phi_\eta(z)$ is

$$
\begin{aligned}
&\phi_\eta(z)\left(1-\alpha\begin{bmatrix}0 & 0 & 1\end{bmatrix}\mathbf{M}'(z^{-1})(\mathbf{M}(z)\mathbf{M}'(z^{-1}))^{-1}\mathbf{M}(z)\begin{bmatrix}0\\0\\1\end{bmatrix}\right) \\
=&\frac{1}{1-\rho z} - \alpha\sum_{k=1}^{d}\frac{\begin{bmatrix}0 & 0 & \phi_\eta(\lambda_k)\end{bmatrix}\mathbf{G}(\lambda_k)}{(z-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)}\frac{\mathbf{K}(z)}{\prod_{k=1}^{d}(1-\lambda_i z)}\begin{bmatrix}0\\0\\1\end{bmatrix}
\end{aligned}
$$

Step 3 First note that[10]

$$\begin{bmatrix}0 & 0 & 1\end{bmatrix}\mathbf{M}'(z^{-1}) = \begin{bmatrix}\frac{z}{z-\rho} & \frac{a(z^{-1})}{b(z^{-1})}\end{bmatrix}, \quad \mathbf{M}(z)\begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{1-\rho z}\\ \frac{a(z)}{b(z)}\end{bmatrix},$$

and

$$\begin{aligned}&(\mathbf{M}(z)\mathbf{M}'(z^{-1}))^{-1}\\ =&\frac{(1-\rho z)(z-\rho)b(z)b(z^{-1})}{(1-\rho z)(z-\rho)b(z)b(z^{-1})+a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1})}\begin{bmatrix}1+\frac{a(z)a(z^{-1})}{b(z)b(z^{-1})} & -\frac{a(z^{-1})}{(1-\rho z)b(z^{-1})}\\ -\frac{za(z)}{(z-\rho)b(z)} & 1+\frac{z}{(z-\rho)(1-\rho z)}\end{bmatrix}.\end{aligned}$$

It is then straightforward to show that

$$\begin{aligned}&1-\alpha\begin{bmatrix}0 & 0 & 1\end{bmatrix}\mathbf{M}'(z^{-1})(\mathbf{M}(z)\mathbf{M}'(z^{-1}))^{-1}\mathbf{M}(z)\begin{bmatrix}0\\0\\1\end{bmatrix}\\ =&\frac{(1-\rho z)(z-\rho)b(z)b(z^{-1})+a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1})-\alpha(a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1}))}{(1-\rho z)(z-\rho)b(z)b(z^{-1})+a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1})}\\ =&\frac{p(z)}{c\prod_{k=1}^{d}(1-\lambda_i z)(z-\lambda_i)}\end{aligned}$$

where $p(z)\equiv(1-\rho z)(z-\rho)b(z)b(z^{-1})+a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1})-\alpha(a(z)a(z^{-1})(z-\rho)(1-\rho z)+zb(z)b(z^{-1}))$ and the denominator follows from equation (A.17) and (A.18).

Now we can derive $\phi_\eta(z)$ as

$$\phi_\eta(z)=\frac{c\prod_{k=1}^{d}(1-\lambda_i z)(z-\lambda_i)}{p(z)}\left(\frac{1}{1-\rho z}-\alpha\sum_{k=1}^{d}\frac{\begin{bmatrix}0 & 0 & \phi_\eta(\lambda_k)\end{bmatrix}\mathbf{G}(\lambda_k)}{(z-\lambda_k)\prod_{\tau\neq k}(\lambda_k-\lambda_\tau)}\frac{\mathbf{K}(z)}{\prod_{k=1}^{d}(1-\lambda_i z)}\begin{bmatrix}0\\0\\1\end{bmatrix}\right)$$

Note that neither λ_i or λ_i^{-1} can be the poles of $\phi_\eta(z)$. The poles of $\phi_\eta(z)$ can only be the roots of $p(z)$ or ρ^{-1} from $\frac{1}{1-\rho z}$. In order to have $\phi_\eta(z)=\psi(z)$, it must be the case that the roots of $b(z)$ are the poles of $\phi_\eta(z)$. Note that when $\alpha=1$,

$$p(z)=(1-\rho z)(z-\rho)b(z)b(z^{-1})).$$

[10] $\mathbf{M}(z)\mathbf{M}'(z^{-1})=\begin{bmatrix}1+\frac{z}{(z-\rho)(1-\rho z)} & \frac{a(z^{-1})}{(1-\rho z)b(z^{-1})}\\ \frac{za(z)}{(z-\rho)b(z)} & 1+\frac{a(z)a(z^{-1})}{b(z)b(z^{-1})}\end{bmatrix}$.

Therefore, if $\alpha \neq 1$, it cannot be the case that $\phi_\eta(z) = \varphi(z)$. $\square$