Rational Expectations Models with Higher Order Beliefs*

Zhen Huo
University of Minnesota
Federal Reserve Bank of Minneapolis

Naoki Takayama
Cabinet Office, Government of Japan

Sunday 30th November, 2014

Please check http://zhenhuo.weebly.com/research.html for the newest version.

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. For this case, we develop a tractable algorithm that can approximate the solution arbitrarily well. The key innovation in our method is to use the factorization identity and Wiener filter to solve signal extraction problems conditional on infinite observables. This method can be used in a wide range of applications. We demonstrate its strong practicability by solving several classical models featuring higher order beliefs, and also a full-blown business cycle model that is driven by confidence shocks.

Keywords: Higher order beliefs, Infinite regress problem, Dispersed Information, Wiener filter.
JEL classifications: E20, E32, F44

*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, 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 his 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 an ARMA process, the equilibrium policy rule always allows a finite-state representation. To make sure signals follow an ARMA process, we start from the case in which the information process is given exogenously. Like in standard problems with symmetric information, solving for the equilibrium requires finding the fixed point in the functional space. Unlike in standard models, when higher order beliefs are involved, it is difficult to figure out 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

[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 (2013), and so on.

the problem which solves for a lag polynomial into a simpler problem which solves for an analytical function, labelled as the frequency-domain method. When signals follow an ARMA process, we prove that the equilibrium policy rule, the lag polynomial, is also of the ARMA form. Therefore, we can find a finite-state representation for the equilibrium policy rule.

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. The restriction that there has to be the same number of signals as shocks is quite limited. In general signal extraction problems, there are more shocks than signals, as criticized in Nimark (2011). This restriction 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 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 extensively in Sargent (1987). In this paper, we extend the application to multivariate case.

We illustrate our method in various applications. We first consider a two-player model in which

[2]See Rondina and Walker (2013), Kasa, Walker, and Whiteman (2014) and Acharya (2013) for example. Walker (2007) and Rondina (2008) solved a special signal process with more shocks than signals.

[3]In Rondina and Walker (2013), they assume a non-invertible shock process.

asymmetric information and strategic complementarity make higher order beliefs relevant.[4] We discuss the case in which agents only receive private signals regarding the economic fundamental (similar to Woodford (2003)), and the case in which agents also receive a public signal regarding the economic fundamental (similar to Angeletos and La'O (2010)). In both cases, we obtain a sharp analytic solution which can be characterized by finite state variables. 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 indeed follows an ARMA process. We also consider a model where agents are randomly matched, an extension of Angeletos and La'O (2013) with persistent shocks. In this case, an agent randomly interacts with a different agent every period, and needs to form higher order beliefs on each of them. Even though it complicates the inference problem, our method is general enough to solve these models as well.

The above first result is for the cases where agents solve their inference problem given an exogenous ARMA signal process. We label them as problems with exogenous information. 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, similar to the concept of information equilibrium defined in Rondina and Walker (2013).

Our second main result is that we prove that in our model with endogenous information, the equilibrium cannot be represented by finite state variables.[5] The endogenous variable that plays an information role 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 noted that it is not because of the infinite regress problem 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 an ARMA process, the individual policy rule will also follow an ARMA process and permit a finite state variable representation. If the endogenous variable does not follow an ARMA process, the signal received by agents cannot follow an 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 this 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

[4]The two-player game should not be taken literally. The two players can be an individual agent and the whole economy.

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

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. We can prove that as the order of the ARMA process increases, it can approximate the true solution arbitrarily well, and we also find that a relatively low order ARMA process can give accurate approximation. 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. The prediction problem still cannot be solved by the Kalman filter. Using our method, each individual's policy rule is solved exactly.

To demonstrate that our method can be applied in an empirically relevant environment, we solve a full-blown business cycle model in a companion paper (Huo and Takayama, 2014). In this paper, the confidence shock is the sole driving force of business cycles, and agents face a complicated learning problem, i.e., they need to forecast the forecasts of others. Different from the applications solved in this paper, agents also make dynamic decisions (the investment decision), and the infinite regress problem becomes much more involved. We show that there is a hump-shaped relationship between the variance of output and the variance of the confidence shock, and under our favored calibration of information frictions, the model with confidence shocks can account for a number of salient features of business cycles.

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 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). 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 (2011). The idea is that 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 as shown in our examples, 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 (2013), 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, a restriction that prevents this method from being applied in more general settings. Furthermore, in previous literature, agents solve a pure forecasting problem most of the time. This paper eliminates these restrictions and 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), 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, 2014), 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 theorems. We show how to jointly use the Kalman filter and the Wiener-Hopf prediction formula to form the optimal expectation with infinite observables. We also show how to obtain a finite-state representation for a rational expectations model with higher order beliefs. Section 4 solves the two-player game with and without public signals. 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 considers the case where an agent has to form higher order beliefs of many different agents. Section 7 discusses an application of the method in a quantitative business cycle model. Section 8 concludes.

2 A Two-Player Model with Infinite Regress Problem

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.

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. The payoff also depends on the action of the other agent and we consider the case with strategic complementarity. However, 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

$$y_{it} = \mathbb{E}[\xi_t|\Omega_{it}] + \alpha\mathbb{E}[y_{jt}|\Omega_{it}],\text{[6]} \tag{2.1}$$

where $\alpha \in (0, 1)$ determines the strength of strategic complementarity and Ω_{it} denotes the information set of agent i at time t. Agent j follows the same strategy. 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

$$\xi_t = \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)}\,\eta_t, \tag{2.2}$$

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

$$x_{it}^1 = \xi_t + \epsilon_{it}, \tag{2.3}$$

$$x_{it}^2 = \xi_t + u_{it}, \tag{2.4}$$

where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $u_{it} \sim N(0, \sigma_u^2)$. Note that the idiosyncratic noises are indexed by i. More compactly, the signal process can be expressed as

$$x_{it} \equiv \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)} \\ 0 & 1 & \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)} \end{bmatrix} \begin{bmatrix} \epsilon_{it} \\ u_{it} \\ \eta_t \end{bmatrix} \equiv M(L)s_{it}, \tag{2.5}$$

The information set of agent i at time t contains all the signals he has received up to time t

$$\Omega_{it} = \left\{ x_{it}^1,\ x_{it}^2,\ x_{it-1}^1,\ x_{it-1}^2,\ x_{it-2}^1,\ x_{it-2}^2,\ \ldots \right\}. \tag{2.6}$$

[6] Here, $y_{it} \in \mathbb{R}$ and $y_{jt} \in \mathbb{R}$. The operator $\mathbb{E}$ denotes the linear projection on the information set.

Agent j receives signals of ξ_t, but are corrupted by his idiosyncratic noises ϵ_{jt} and u_{jt}. 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 $\mathbb{E}[\,\cdot \mid \Omega_{it}]$ from now on.

Remark Several remarks about the model should be made here before we move on.

1. A wide range of models can be interpreted as the two-player model. If we assume that there are a continuum of agents in the economy, and each individual agent i interacts with the economy average $y_t = \int y_{jt}$, the model becomes

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_t]. \tag{2.7}$$

   As we show in Section 4, the solution to this model remains the same as the original model (2.1). What matters is whether to infer the action of a fixed agent (Section 4), or to infer the action of a random agent that changes over time (Section 6).

2. To introduce the infinite regress problem, it will be sufficient if agents only receive one of the two signals. The assumption that agents receive multiple signals is to demonstrate that our method can manage multivariate systems.

3. The information structure we have specified in equation (2.5) is a very special one. We can relax this assumption to allow any finite number of signals that follows any finite ARMA process. The structure we adopt here should not be taken in a narrow way. For example, we allow some of the signals to be shared by all agents (Section 4.2), and allow some of the signals to contain endogenous information (Section 5).

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,

$$y_{jt} = \mathbb{E}_{jt}[\xi_t] + \alpha \mathbb{E}_{jt}[y_{it}]. \tag{2.8}$$

We can repeatedly substitute equation (2.8) into equation (2.1), and vice versa, which leads to

$$\begin{aligned}
y_{it} &= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_{jt}] \\
&= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\left[\mathbb{E}_{jt}[\xi_t] + \alpha \mathbb{E}_{jt}[y_{it}]\right] \\
&= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\mathbb{E}_{jt}[\xi_t] + \alpha^2 \mathbb{E}_{it}\mathbb{E}_{jt}[y_{it}] \\
&= \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}\mathbb{E}_{jt}[\xi_t] + \alpha^2 \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}[\xi_t] + \alpha^3 \mathbb{E}_{it}\mathbb{E}_{jt}\mathbb{E}_{it}[y_{jt}] \\
&\ \ \vdots \\
&= \sum_{k=0}^{\infty} \alpha^k \ \mathbb{E}_{it}^{k+1}[\xi_t], \qquad (2.9)
\end{aligned}$$

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. However, this approach does not work in general, due to a growing complexity with the order of beliefs.

Similarly, if we consider model (2.7), successive substitution leads to

$$y_{it} = \sum_{k=0}^{\infty} \alpha^k \ \mathbb{E}_{it}\overline{\mathbb{E}}_t^k[\xi_t]. \qquad (2.10)$$

Here, as opposed to inferring agent j's beliefs, the higher order beliefs $\overline{\mathbb{E}}_t^k[\xi_t]$ are about the economy average expectations of ξ_t, defined recursively by

$$\begin{aligned}
\overline{\mathbb{E}}_t^0[\xi_t] &= \xi_t \\
\overline{\mathbb{E}}_t^1[\xi_t] &= \int \mathbb{E}_{jt}[\xi_t] \\
\overline{\mathbb{E}}_t^k[\xi_t] &= \int \mathbb{E}_{jt}\overline{\mathbb{E}}_t^{k-1}[\xi_t].
\end{aligned}$$

In both cases, it is apparent that agents' optimal response is related to infinite higher order 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.

2.3 Equilibrium

Recall that the information set of agent i is $\Omega_{it} = x_i^t$. The linear policy rule of agent i belongs to the space spanned by square-summable linear combinations of current and past realizations of x_{it}. We use $\mathcal{H}_t^x$ to denote this space. We assume that the policy rule takes the following form

$$y_{it} = \sum_{k=0}^{\infty} h_{1k} x_{it-k}^1 + \sum_{k=0}^{\infty} h_{2k} x_{it-k}^2, \tag{2.11}$$

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, there is no way 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 we have to assume it is necessary to keep track of the entire history of signals.

More compactly, we use lag polynomials to denote the infinite sum

$$y_{it} = h_1(L) x_{it}^1 + h_2(L) x_{it}^2, \tag{2.12}$$

with $h_1(L) = \sum_{k=0}^{\infty} h_{1k} L^k$ and $h_2(L) = \sum_{k=0}^{\infty} h_{2k} L^k$.

To make sure that y_{it} is co-variance stationary, the infinite sequences $\{h_{1\tau}\}_{\tau=0}^{\infty}$ and $\{h_{2\tau}\}_{\tau=0}^{\infty}$ have to be in the square-summable space ℓ^2.[7] From now on, if an infinite sequence $\phi = \{\phi_k\}_{k=0}^{\infty} \in \ell^2$, then we denote $\phi(L) = \sum_{k=0}^{\infty} \phi_k L^k$ as its corresponding lag polynomial. The definition of the equilibrium is straightforward.

Definition 2.1 (Signal form)**.** *Given the signal process (2.5), the equilibrium of model (2.1) is a policy rule $h = \{h_1, h_2\} \in \ell^2 \times \ell^2$, such that*

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

[7] Hansen and Sargent (1980), Kasa (2000) assume that the policy rule ϕ belong to β-summable space, i.e., $\sum_{k=0}^{\infty} \beta^k \phi_k^2 < \infty$. This is a less strict requirement than original ℓ^2 assumption, which arises naturally in linear-quadratic type models. However, it is less obvious whether this relaxation is valid or not in our model setting, and we will work with the original ℓ^2 space in this paper.

where

$$y_{it} = h_1(L)x^1_{it} + h_2(L)x^2_{it},$$
$$y_{jt} = h_1(L)x^1_{jt} + h_2(L)x^2_{jt}.$$

Since the signals $\{x_{it}\}$ are ultimately generated by the underlying shocks $\{s_{it}\}$, y_{it} also lies in the space spanned by the square-summable linear combinations of current and past shocks, denoted by $\mathcal{H}^s_t$. It should be clear that $\mathcal{H}^x_t \subset \mathcal{H}^s_t$. We say that the equilibrium is of signal form if the equilibrium policy is written in terms of signals, and the equilibrium is of innovation form if it is written in terms of the underlying shocks. The equilibrium in innovation form is defined as follows

Definition 2.2 (Innovation form)**.** *Given the signal process (2.5), the equilibrium of model (2.1) is a policy rule $\phi = \{\phi_1, \phi_2, \phi_3\} \in \ell^2 \times \ell^2 \times \ell^2$, such that*

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

where

$$y_{it} = \phi_1(L)\epsilon_{it} + \phi_2(L)u_{it} + \phi_3(L)\eta_t,$$
$$y_{jt} = \phi_1(L)\epsilon_{jt} + \phi_2(L)u_{jt} + \phi_3(L)\eta_t.$$

In the literature, when solving the infinite regress problem in the frequency domain, the innovation form is exclusively used. The advantage of working with innovation form is that all the objects are expressed in terms of the underlying shocks and it is convenient to discuss its statistical properties. However, from an economic perspective, it is more natural to think of the policy rule in terms of signals, because agents do not observe those shocks directly.[8] In Theorem 4, we show that there is a one-to-one mapping between the equilibrium in signal form and in innovation form.

In terms of the existence and uniqueness of the equilibrium, we have the following result.

Proposition 2.1. *Assume that the signals follow a co-variance stationary process. If $\alpha \in (0, 1)$, then there exists a unique equilibrium of model (2.1).*

Proof. See Appendix A.1 for proof. □

The core of the proof is to show that the equilibrium is a fixed point of a contraction mapping. On one hand, to prove this proposition, we only require that the signals follow a co-variance stationary

[8] Kasa (2000) claims that the limited-information equilibrium does not exist in the space spanned by signals but only exists in the space spanned by the innovations. We find this conclusion questionable.

process, but not necessarily a finite ARMA process. On the other hand, this proposition does not imply whether the policy rule in equilibrium permits a finite-state representation or not. In principle, it could be that agents do need to keep track of the entire history of observables. Next theorem, however, shows that the equilibrium indeed has a finite-state representation when the signals follow a finite ARMA process.

2.4 Finite-state representation

Theorem 1. *Assume that (1) the exogenous variable ξ_t follows*

$$\xi_t = \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)}\eta_t.$$

(2) The signals follow the following co-variance stationary process (2.5)

$$x_{it} = \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)} \\ & & \\ 0 & 1 & \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)} \end{bmatrix} \begin{bmatrix} \epsilon_{it} \\ u_{it} \\ \eta_t \end{bmatrix}.$$

(3) The structural parameter $\alpha \in (0,1)$.

Then there exists a unique solution $y_{it} = h_1(L)x_{it}^1 + h_2(L)x_{it}^2$ satisfies model (2.1)

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

The equilibrium policy rule $h_1(L)$ and $h_2(L)$ have the following properties

1. *Both $h_1(L)$ and $h_2(L)$ have a finite ARMA representation*

$$y_{it} = \begin{bmatrix} h_1(L) & h_2(L) \end{bmatrix} \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \begin{bmatrix} \tau_1 \frac{\Pi_{k=1}^{n}(1+\zeta_k^1 L)}{\Pi_{k=1}^{m}(1-\vartheta_k L)} & \tau_2 \frac{\Pi_{k=1}^{n}(1+\zeta_k^2 L)}{\Pi_{k=1}^{m}(1-\vartheta_k L)} \end{bmatrix} \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} \tag{2.13}$$

   where the order of the ARMA process m and n, the coefficients τ_1, τ_2, $\{\vartheta_k\}_{k=1}^m$, $\{\zeta_k^1\}_{k=1}^n$, and $\{\zeta_k^2\}_{k=1}^n$ are all functions of the structural parameter α and the parameters that determine the signal process.

2. *Let $r = max\,\{m, n\}$. Given a particular signal realization $\{x_{it}\}_{t=-\infty}^{-1}$, there exists r state variables $z_{it} = [z_{it}^1, z_{it}^2, \ldots, z_{it}^r]'$, such that the policy rule in (2.13) has the following finite-state representation*

$$y_{it} = \Gamma_x\, x_{it} + \Gamma_z\, z_{it}, \tag{2.14}$$

with the law of motion of z_{it}

$$z_{it+1} = \Upsilon_x x_{it} + \Upsilon_z z_{it} \tag{2.15}$$

The initial state z_{i0} is given by

$$z_{i0} = (I_r - \Upsilon_z L)^{-1} \Upsilon_x x_{i-1} \tag{2.16}$$

The constant matrices $\Gamma_x, \Gamma_z, \Upsilon_x$, and Υ_z are all functions of τ_1, τ_2, $\{\vartheta_k\}_{k=1}^m$, and $\{\zeta_k^1\}_{k=1}^n$ in equation (2.13).

Proof. The proof of this theorem is a subset of the proof of Theorem 2, and the exact form of equation (2.13) can be derived by Theorem 3 in the next section. □

The first part of this theorem establishes that the equilibrium policy rule follows a finite ARMA process in terms of the signals. The second part of this theorem states that the policy rule has a finite-state representation, which is a natural result of the first part. Therefore, there indeed exists a small set of state variables that are sufficient for agents' inference problem. This theorem also implies that the infinite sum of higher order beliefs in equation (2.9) follows a finite ARMA process, even though $\mathbb{E}_{it}^k[\xi_t]$ follows an infinite ARMA process as k approaches to infinity.

To solve for the equilibrium policy rules $h_1(L)$ and $h_2(L)$, the difficulty lies in how to solve the inference problem

$$\mathbb{E}_{it}[y_{jt}] = \mathbb{E}_{it}[h_1(L)x_{jt}^1 + h_2(L)x_{jt}^2],$$

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). A key step to employ the Wiener is to find the Wold representation of the signal process, which is not provided by the Wiener filter itself but can be obtained by the Kalman filter. Therefore, a joint use of the Kalman filter and the Wiener filter solves this inference problem. The lack of an efficient way to find the Wold representation is exactly what prevents others from solving models with higher order beliefs, and we show that the Kalman filter can achieve this goal with ease. After solving $\mathbb{E}_{it}[y_{jt}]$, it turns out that $h_1(L)$ and $h_2(L)$ are of finite ARMA type, and it allows a finite-state representation.

The model we considered in this section is a very special one in the following sense: (1) there is only one choice variable y_{it}; (2) there is no endogenous state variables, such as capital; (3) there is no need to forecast variables in the future; (4) the signal process is very special. These limitations make model (2.1) only theoretically interesting, and far from empirically relevant. In the following section, we eliminate these restrictions, and extend Theorem 1 to a much more general statement.

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, and state the main theorem that the equilibrium policies admit finite-state representation. We then show how to prove this theorem in steps. The key part is to use the Wold representation and the Wiener filter to solve the general signal extraction problem.

3.1 General rational expectations models

Now we move to the general form of the linear system. The input of the model includes two parts: the first part is the signal process; the second part is the linear system which corresponds to the equilibrium conditions that various kinds of variables need to satisfy. There are three kinds of variables involved here: choice variables, choice variables chosen by others, and exogenous variables.

Signal process Assume that the signals observed by agents follow a finite ARMA process,

$$x_t = \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^n \end{bmatrix} = \begin{bmatrix} \frac{a_{11}(L)}{b_{11}(L)} & \cdots & \frac{a_{1m}(L)}{b_{1m}(L)} \\ \vdots & \ddots & \vdots \\ \frac{a_{n1}(L)}{b_{n1}(L)} & \cdots & \frac{a_{nm}(L)}{b_{nm}(L)} \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} = M(L)s_t, \tag{3.1}$$

where the signal x_t is a stochastic $n \times 1$ vector and the shock s_t is a stochastic $m \times 1$ vector. We allow m to be different from n. We normalize the co-variance matrix of s_t to be an identity matrix. In each element of $M(L)$, $a_{ij}(L)$ and $b_{ij}(L)$ are finite order polynomials in the lag operator L. Particularly,

$$a_{ij}(L) = \sum_{k=0}^{q_{ij}} \alpha_{ijk} L^k,$$
$$b_{ij}(L) = \sum_{k=0}^{p_{ij}} \beta_{ijk} L^k,$$

and we normalize $\beta_{ij0} = 1$. The information set is $\Omega_t = x^t = \{x_t, x_{t-1}, x_{t-2}, \dots\}$.

Choice variable We assume there are d choice variables, which are functions of the signals:

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{dt} \end{bmatrix} = h(L)x_t = \begin{bmatrix} h_{11}(L) & \dots & h_{1n}(L) \\ \vdots & \dots & \vdots \\ h_{d1}(L) & \dots & h_{dn}(L) \end{bmatrix} \begin{bmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{bmatrix} = h(L)M(L)s_t. \tag{3.2}$$

$h(L)$ is the equilibrium policy rule we want to solve. We assume that each element in $h(L)$ has an infinite MA representation. We do not impose that $h(L)$ admits a finite ARMA representation in the first place (even though we prove this is indeed the case later). Because these choice variables only depend on signals up to t, $h_{ij}(L)$ cannot contain any negative powers in L. To write it more compactly for future derivation, define

$$\phi(L) \equiv \begin{bmatrix} h_{11}(L) & \dots & h_{1n}(L) & \dots & h_{d1}(L) & \dots & h_{dn}(L) \end{bmatrix}. \tag{3.3}$$

$\phi(L)$ effectively collapse all the lag polynomials to be solved into a vector, the dimension of which is denoted as $w \equiv dn$. Reversely, the elements of y_t can be expressed in terms of $\phi(L)$ as

$$\begin{aligned} y_{it} &= \phi(L)A_i x_t & (3.4) \\ &= \phi(L)A_i M(L)s_t & (3.5) \end{aligned}$$

where A_i is the constant matrix that selects $[h_{i1} \dots h_{in}]$ from $\phi(L)$. Later we will use $h(L)$ and $\phi(L)$ interchangeably.

Endogenous variables related to other agents' actions Crucially, the optimal policy may depend on other agents' actions or depend on some aggregate endogenous variables. These variables cannot be observed, but matter for agents' payoff. Assume there are d_f such endogenous variables, denoted by $f_t = [f_{it}, \dots, f_{d_f t}]'$ denote these endogenous variables. They are related to the policy rule $\phi(L)$ and the underlying shocks s_t in the following way

$$f_{it} = \phi(L)f^i(L)s_t = \phi(L) \begin{bmatrix} f^i_{11}(L) & \dots & f^i_{1m}(L) \\ \vdots & \dots & \vdots \\ f^i_{w1}(L) & \dots & f^i_{wm}(L) \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} \tag{3.6}$$

Here, each $f^i(L)$ is a $w \times m$ matrix in the lag operator L. We assume that all the elements in $f^i(L)$ are finite rational functions in L and do not contain negative powers of L in expansion (others' action cannot be a function of future shocks either).

Note that actions of others may also depend on shocks other than $\{s_{1t}, \dots, s_{mt}\}$. However, these shocks are uncorrelated with the shocks $\{s_{1t}, \dots, s_{mt}\}$ that drive $\{x_t\}$, and the best forecasts of those shocks conditional on $\{x_t\}$ are zero. As a result, what is relevant for agents are the parts that are correlated with $\{s_{1t}, \ldots, s_{mt}\}$.

Exogenous variables Generally, the optimal policy depends on the evolution of some exogenous variables. We assume there are d_g such variables, denoted by $g_t = [g_{it}, \ldots, g_{d_g t}]'$

$$g_t = g(L)s_t = \begin{bmatrix} g_{11}(L) & \ldots & g_{1m}(L) \\ \vdots & \ldots & \vdots \\ g_{d_g 1}(L) & \ldots & g_{d_g m}(L) \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} \tag{3.7}$$

Note that these exogenous variables are independent of the equilibrium policy rule $\phi(L)$ Similarly, we assume that all the elements of $g(L)$ are rational functions in L.

General model Assume the policy rule needs to satisfy the following linear system in equilibrium

$$\sum_{j=0}^{p} C^{y,j} L^j y_t + \sum_{j=0}^{p} \mathbb{E}\left[C^{f,j} L^j f_t + C^{g,j} L^j g_t \middle| x^t \right] + \sum_{j=1}^{q} \mathbb{E}\left[C^{y,-j} L^{-j} y_t + C^{f,-j} L^{-j} f_t + C^{g,-j} L^{-j} g_t \middle| x^t \right] = 0 \tag{3.8}$$

For $j \in \{-q, \ldots, p\}$, $C^{y,j}$ is a constant $d \times d$ matrix, $C^{f,j}$ is a constant $d \times d_f$ matrix, and $C^{g,j}$ is a constant $d \times d_g$ matrix. These matrices are structural parameters that result from optimality conditions and resource constraints. This system of equations incorporates the possibilities that the choice variables y_t depend on the past, the current and the future values of the endogenous variables of others and the exogenous variables, and also y_t's own and future values. This specification includes the majority of applications that one may encounter.

Special cases The structure we have specified includes two special cases which are common in the literature.

1. Perfect information.

$$\sum_{j=0}^{p} C^{y,j} L^j y_t + \sum_{j=0}^{p} \mathbb{E}\left[C^{f,j} L^j f_t + C^{g,j} L^j g_t \middle| s^t \right] + \sum_{j=1}^{q} \mathbb{E}\left[C^{y,-j} L^{-j} y_t + C^{f,-j} L^{-j} f_t + C^{g,-j} L^{-j} g_t \middle| s^t \right] = 0$$

In standard real business cycle models and New Keynesian models without information frictions, the underlying shocks $\{s_t\}$ are observed directly by agents. That is, the space spanned by shocks is the same as the space spanned by signals, $\mathcal{H}_t^s = H_t^x$. Also, because all the shocks are observed directly, the actions of other agents are also known perfectly. As a result, the expectations in model (3.8) can be calculated in a trivial way.

2. Imperfect information, but no roles of higher order beliefs [9]

$$\sum_{j=0}^{p} C^{y,j} L^j y_t + \sum_{j=0}^{p} \mathbb{E}\left[C^{g,j} L^j g_t \middle| x^t\right] + \sum_{j=1}^{q} \mathbb{E}\left[C^{y,-j} L^{-j} y_t + C^{g,-j} L^{-j} g_t \middle| x^t\right] = 0$$

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

The solution to model (3.8) defined as follows

Definition 3.1. *Given the signal process (3.1), a solution to model (3.8) (or an equilibrium) is a matrix of lag polynomials $h(L)$ or equivalently $\phi(L)$, such that*

1. *For all (i, j), $h_{ij}(L)$ has an infinite MA representation*

$$h_{ij}(L) = \sum_{k=0}^{\infty} h_{ijk} L^k,$$

with $\sum_{k=0}^{\infty} h_{ijk} < \infty$.

2. *For all possible realizations of $\{x_t\}$,*

$$y_t = h(L)x_t$$

satisfies equation (3.8).

Given the model, we are interested in the following questions:

1. Under what conditions does a unique solution to this problem exist?
2. Suppose there indeed exists a $h(L)$ that solves the problem, what its formula?
3. Does the solution admit a finite-state representation which allows agents to summarize the past information using a small set of statistics?

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

Theorem 3, which involves more technical details, answers the first two questions. The following theorem answers the third question.

Theorem 2 (Finite-state representation)**.** *Suppose the signal process follows (3.1) and the model (3.8) has a solution $y_t = h(L)x_t$. Then $y_t = h(L)x_t$ has a finite ARMA representation*

$$y_t = h(L)x_t = \begin{bmatrix} \frac{c_{11}(L)}{d_{11}(L)} & \cdots & \frac{c_{1n}(L)}{d_{1n}(L)} \\ \vdots & \ddots & \vdots \\ \frac{c_{d1}(L)}{d_{d1}(L)} & \cdots & \frac{c_{dn}(L)}{d_{dn}(L)} \end{bmatrix} \begin{bmatrix} x_{1t} \\ \vdots \\ x_{nt} \end{bmatrix}, \tag{3.9}$$

where $c_{ij}(L)$ and $d_{ij}(L)$ are finite degree polynomials in the lag operator L.

Given a particular signal realization $\{x_t\}_{t=-\infty}^{-1}$, there exists a finite set of state variables z_t, such that

$$y_t = \Gamma_x \, x_t + \Gamma_z \, z_t, \tag{3.10}$$

with the law of motion of z_t

$$z_{t+1} = \Upsilon_x x_t + \Upsilon_z z_t. \tag{3.11}$$

The initial state z_0 is given by

$$z_0 = (I - \Upsilon_z L)^{-1} \, \Upsilon_x x_{-1} \tag{3.12}$$

Proof. See Appendix A.10 for proof. □

This theorem implies that 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, given that the signal process is of ARMA type. We present the proof of this theorem and the proof of Theorem 3 in steps in the following subsections.

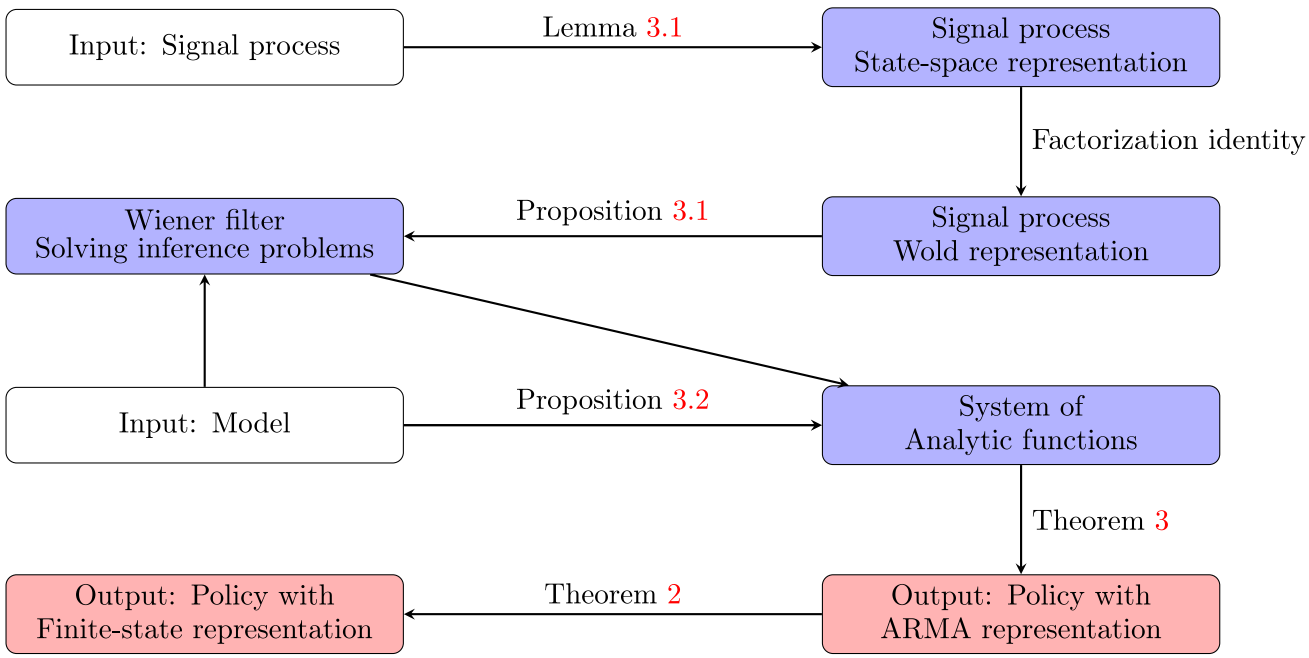
3.2 Preview of the main steps

The proof of these theorems is quite lengthy and it involves a number of building blocks. The initial input includes the signal process (3.1) and the model (3.8). Here, we first sketch the main steps that lead to Theorem 3 and Theorem 2, which is also shown in Figure 1 .

Step 1: Given the signal process (3.1), find its state-space representation.

Step 2: With the state-space of the signal process, use the innovation representation and factorization identity matrix to find the Wold representation of the signal process.

Figure 1: Main Steps of Solving Rational Expectations Models with Higher Order Beliefs



Step 3: With the Wold representation of the signal process, use Wiener filter to solve the inference problem in model (3.8).

Step 4: Applying the Riesz-Fisher Theorem, 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.

Step 5: Use Cramer's rule to solve the system of analytic functions, which leads to the solution $h(L)$ with ARMA representation.

Step 6: Given the solution with ARMA representation, find its finite-state representation.

3.3 Mathematical background: z transformation

By the Riesz-Fisher Theorem, there is a one-to-one mapping between the space of square-summable sequences and the space of complex-valued functions. Given a two-sided lag polynomials

$$\psi(L) = \sum_{k=-\infty}^{\infty} \psi_k L^k,$$

with $\sum_{k=-\infty}^{\infty} |\psi_k|^2 < \infty$, we will use the complex-valued function $\psi(z)$ to denote its corresponding z transformation

$$\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k,$$

where $\psi(z)$ is defined on the unit circle.

If $\psi(L)$ is a one-sided polynomial with $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$, then its z transformation is an analytic function on the open unit disk.

Particularly, assume there are two univariate co-variance stationary processes

$$x_t = M(L)s_t,$$
$$y_t = \psi(L)s_t.$$

The auto-covariance generating function for x_t is

$$\rho_{xx}(z) = M(z)M'(z^{-1}),$$

and the cross-covariance generating function between y_t and x_t is

$$\rho_{yx}(z) = \psi(z)M'(z^{-1}).$$

Most of the time, working with a complex function is much more convenient than working with a square-summable sequence.

3.4 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 s^t contain more information than the signals, and the prediction conditional on s^t is different from the prediction conditional on x^t. The Wold representation provides a new sequence of shocks w^t. Different from the underlying shocks s^t, the space spanned by the signals x^t is equivalent to the space spanned by w^t, and we can conduct the linear projection on w^t. Given a finite ARMA signal process, in this subsection we present how to find its state-space representation and Wold representation using the factorization identity.

Lemma 3.1. *Assume that x_t follows a finite ARMA process and is co-variance stationary,*

$$x_t = \begin{bmatrix} x_t^1 \\ \vdots \\ x_t^n \end{bmatrix} = \begin{bmatrix} \frac{a_{11}(L)}{b_{11}(L)} & \cdots & \frac{a_{1m}(L)}{b_{1m}(L)} \\ \vdots & \ddots & \vdots \\ \frac{a_{n1}(L)}{b_{n1}(L)} & \cdots & \frac{a_{nm}(L)}{b_{nm}(L)} \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} = M(L)s_t, \tag{3.13}$$

where x_t is a $n \times 1$ vector and s_t is a $m \times 1$ vector. The co-variance matrix of s_t is normalized to be an identity matrix. In each element of $M(L)$, $a_{ij}(L)$ and $b_{ij}(L)$ are finite degree polynomials in the lag operator L. Particularly,

$$a_{ij}(L) = \sum_{k=0}^{q_{ij}} \alpha_{ijk} L^k$$
$$b_{ij}(L) = \sum_{k=0}^{p_{ij}} \beta_{ijk} L^k$$

and we normalize $\beta_{ij0} = 1$. The signal process admits at least one state-space representation. The state equation is

$$Z_t = FZ_{t-1} + Qs_t,$$

and the observation equation is

$$x_t = HZ_t,$$

where F, Q and H are functions of $\left\{p_{ij}, q_{ij}, \{\alpha_{ijk}\}_{k=1}^{q_{ij}}, \{\beta_{ijk}\}_{k=1}^{p_{ij}}\right\}$.

In addition, the eigenvalues of F all lie inside the unit circle.

Proof. See Appendix A.3 for proof. □

This lemma states that any finite ARMA process has a state-space representation. Note that there are many different state-state representations for the same ARMA process. Generally, we can write the state equation as

$$Z_t = FZ_{t-1} + Qs_t,$$

and the observation equation as

$$x_t = HZ_t + Rv_t,$$

where the covariance matrix of v_t is also an identity matrices. Lemma 3.1 only provides one of the state-space representation with the feature that there is no shock in the observation equation.

Finding the state-space representation is a necessary step to find the Wold representation of the signal process. Suppose that we have $B(L)$ and $\{w_t\}$ such that

$$x_t = M(L)s_t = B(L)w_t, \tag{3.14}$$

$B(L)$ is invertible,[10] and w_t is serially uncorrelated shocks with co-variance matrix V, then we say $x_t = B(L)w_t$ is a Wold representation of x^t. Since $B(L)$ is invertible, x^t contains the same information as w^t, i.e., $\mathcal{H}_t^x = \mathcal{H}_t^w$. Further, equation (3.14) implies that

$$\rho_{xx}(z) = M(z)M'(z^{-1}) = B(z)VB'(z^{-1}). \tag{3.15}$$

$B(z)$ and V is called a canonical factorization of $\rho_{xx}(z)$. Therefore, find the Wold representation is equivalent to find the canonical factorization. The following theorem provides the canonical factorization for the state-space representation of the signal process x_t, which uses the factorization identity.

Theorem (Canonical Factorization)**.** *Let F denote an $(r \times r)$ matrix whose eigenvalues are all inside the unit circle; let $Q'Q$ or $R'R$ be positive definite matrix of dimension $(r \times r)$ or $(n \times n)$; let H denote an arbitrary $(n \times r)$ matrix. Let P satisfy*

$$P = F[P - PH'(HPH' + R'R)^{-1}HP]F' + Q'Q$$

and K be defined as

$$K = PH'(HPH' + R'R)^{-1}$$

Then

1. *The eigenvalues of $(F - FKH)$ are all inside the unit circle.*

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

2. The canonical factorization is

$$\begin{aligned}\rho_{xx}(z) &= H[I_r - Fz]^{-1}Q'Q[I_r - Fz^{-1}]^{-1}H' + R'R \\ &= [I_n + H(I_r - Fz)^{-1}FKz][HPH' + R'R][I_n + K'F'(I_r - F'z^{-1})^{-1}H'z^{-1}] \\ &= B(z)VB'(z^{-1}).\end{aligned}$$

3. $B(z)$ *is*

$$B(z) = I_n + H[I_r - Fz]^{-1}FKz,$$

the inverse of $B(z)$ *is*

$$B(z)^{-1} = I_n - H[I_r - (F - FKH)z]^{-1}FKz,$$

and the co-variance matrix V *is*

$$V = HPH' + R'R$$

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

To prove this theorem, one essentially uses the Kalman filter. The requirement that all the eigenvalues of F lie inside the unit circle guarantees $(I_r - Fz)$ is invertible. The eigenvalues of $(F - FKH)$ are very important in understanding the prediction problem, which essentially determines the persistence of the forecasts.

3.5 Wiener-Hopf prediction formula

Now we turn to the inference problems incorporated in equation (3.8). 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*

$$x_t = M(L)s_t,$$

and y_t *is a univariate co-variance stationary process*

$$y_t = \psi(L)s_t.$$

Assume all the elements of $M(L)$ *and* $\psi(L)$ *have an infinite MA representation. The canonical* *factorization of $\rho_{xx}(z)$ is given by*

$$\rho_{xx}(z) = M(z)M'(z^{-1}) = B(z)VB'(z^{-1}).$$

Then the optimal linear prediction of y_t conditional on $\{x_t\}$ is

$$\mathbb{E}[y_t|x^t] = \Big[\rho_{yx}(L)B'(L^{-1})^{-1}\Big]_+ V^{-1}B(L)^{-1}. \tag{3.16}$$

Proof. See Appendix A.4 for proof. □

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

Lemma 3.2. *Assume the signal process follows equation (3.1). Then*

$$M'(z^{-1})B'(z^{-1})^{-1} = \frac{1}{\Pi_{k=1}^{u}(z-\lambda_k)}G(z) \tag{3.17}$$

where $B(z)$ is given by the Canonical Factorization Theorem , $G(z)$ is a polynomial matrix in z, and $\{\lambda_k\}_{k=1}^{u}$ are non-zero eigenvalues of $F - FKH$ which all lie inside the unit circle.

Proof. See Appendix A.5 for proof □

Proposition 3.1. *Given the signal process in equation (3.1), suppose there is a univariate random variable y_t follows*

$$y_t = \psi(L)s_t,$$

where the elements of $\phi(L)$ has an infinite MA representation.

Assume $\{\lambda_k\}_{k=1}^{u}$ in Lemma 3.2 are distinct, the prediction formula for current and past y_t is

$$\mathbb{E}\left[L^j y_t \mid x^t\right] \;=\; \psi(L)L^jM'(L^{-1})\rho_{xx}(L)^{-1}x_t - \sum_{k=1}^{u}\frac{\psi(\lambda_k)\lambda^kG(\lambda_k)V^{-1}B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}x_t \tag{3.18}$$

where $j = \{0, 1, 2 \ldots\}$.

The prediction formula for j-step ahead prediction is

$$\begin{aligned}\mathbb{E}\left[L^{-j}y_t \mid x^t\right] &= \psi(L)L^{-j}M'(L^{-1})\rho_{xx}(L)^{-1}x_t - \sum_{k=1}^{u}\frac{\psi(\lambda_k)G(\lambda_k)L^{-j}V^{-1}B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}x_t \\ &- \sum_{\ell=0}^{j-1}\ell!\,L^{\ell-j}\left[\frac{\psi(L)G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u}\frac{\psi(\lambda_k)G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1}B(L)^{-1}x_t \end{aligned} \tag{3.19}$$

where $[\cdot]_0^{(\ell)}$ *denote the* ℓ*-th order derivative evaluated at 0.*

Proof. See Appendix A.6 for proof. □

The key in applying the Wiener-Hopf prediction formula is to find the Wold representation for x_t or the canonical factorization for $M(z)$. When the number of signals equals the number of shocks, $M(L)$ is a square matrix. Suppose $M(L)$ is invertible, then $M(L)$ itself is a Wold representation and the Wiener-Hopf prediction formula can be applied directly. This corresponds to the case when there is no information friction or the signals fully reveal the state of the economy. If $M(L)$ is a square but not an invertible matrix, then there exists at least one inside root of the determinant of $M(L)$. In this case, the Wold representation can be found by multiplying the Blaschke matrices $B_j(z)$ to flip the inside roots outside the unit circle

$$B(z) \quad = \quad M(z)\Pi_j(W_jB_j(z)).$$

The details of the Blaschke matrix can be found in Rozanov (1967). Kasa (2000), Rondina and Walker (2013), Kasa, Walker, and Whiteman (2014) and Acharya (2013) all use this method to find the Wold representation.

If the number of shocks is larger than the number signals, $M(L)$ is a non-square matrix and is not invertible. To find the canonical factorization of $M(L)$ is more involved, but we just show this can be achieved by using the Canonical Factorization Theorem.

As criticized by Nimark (2011), in most signal extraction problems, the number of shocks is larger than the number of signals. Existing literature restricts 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. 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 (2013) and Kasa, Walker, and Whiteman (2014) both have square observation matrix, and to prevent the price from fully revealing the information, they have to abandon the standard AR(1) process but assume that the underlying shock follows a confounding process.

More importantly, 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 easily solve the signal extraction problem.

3.6 System of analytic functions

After we apply the Wiener filter, solving for $h(L)$ or $\phi(L)$ in model (3.8) 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. *Given the signal process (3.1), there exists a solution $\phi(L)$ to model (3.8) if and only if there exists a vector analytic function $\phi(z)$ that solves*

$$T(z)\phi(z) = D\left[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\right] \tag{3.20}$$

where $T(z)$ is a $w \times w$ matrix given by

$$T(z) \equiv \begin{bmatrix} \sum_{j=-q}^{p} z^j \left[\sum_{i=1}^{r} C_{1,i}^{y,j} A_i + \sum_{i=1}^{v} C_{1,i}^{f,j} f_i(z) M'(z^{-1}) \rho_{xx}(z)^{-1}\right]' \\ \vdots \\ \sum_{j=-q}^{p} z^j \left[\sum_{i=1}^{r} C_{r,i}^{y,j} A_i + \sum_{i=1}^{v} C_{r,i}^{f,j} f_i(z) M'(z^{-1}) \rho_{xx}(z)^{-1}\right]' \end{bmatrix} \tag{3.21}$$

and $D\left[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\right]$ *is a* $w \times 1$ *vector given by*

$$D\left[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\right] = \tag{3.22}$$

$$\left[\begin{array}{l}
\Bigg\{ -\sum_{j=-q}^{p} C_1^{g,j} g(z) z^j M'(z^{-1}) \rho_{xx}(z)^{-1} \\
+\sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j \left[\sum_{i=1}^{d_f} C_{1,i}^{f,j}\phi(\lambda_k) f_i(\lambda_k) + \sum_{i=1}^{d_g} C_1^{g,j} g(\lambda_k)\right] G(\lambda_k) V^{-1} B(z)^{-1}}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} \\
+\sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, z^{\ell-j} \Bigg(\left[\frac{\sum_{i=1}^{d}\phi(z) C_{1,i}^{y,-j} A_i M(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{d}\phi(\lambda_k) C_{1,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\
\qquad + \left[\frac{\sum_{i=1}^{d_f}\phi(z) C_{1,i}^{f,-j} f_i(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{d_f}\phi(\lambda_k) C_{1,i}^{f,-j} f_i(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\
\qquad + \left[\frac{C_{1,i}^{g,-j} g(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{1,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \Bigg) V^{-1} B(z)^{-1} \Bigg\}' \\
\vdots \\
\vdots \\
\Bigg\{ -\sum_{j=-q}^{p} C_d^{g,j} g(z) z^j M'(z^{-1}) \rho_{xx}(z)^{-1} \\
+\sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j \left[\sum_{i=1}^{d_f} C_{d,i}^{f,j}\phi(\lambda_k) f_i(\lambda_k) + \sum_{i=1}^{d_g} C_d^{g,j} g(\lambda_k)\right] G(\lambda_k) V^{-1} B(z)^{-1}}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} \\
+\sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, z^{\ell-j} \Bigg(\left[\frac{\sum_{i=1}^{d}\phi(z) C_{1,i}^{y,-j} A_i M(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{d}\phi(\lambda_k) C_{r,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\
\qquad + \left[\frac{\sum_{i=1}^{d_f}\phi(z) C_{d,i}^{f,-j} f_i(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{d_f}\phi(\lambda_k) C_{d,i}^{f,-j} f_i(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\
\qquad + \left[\frac{C_{d,i}^{g,-j} g(z) G(z)}{\Pi_{k=1}^{u}(z-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{d,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \Bigg) V^{-1} B(z)^{-1} \Bigg\}'
\end{array}\right]$$

Proof. See Appendix A.7 for proof. □

To solve for $\phi(z)$, one can use the Cramer's rule. However, one also needs to determine the following constants, $\{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}$, which are generated when applying the Wiener-Hopf prediction formula,. As discussed in Whiteman (1983), these constants can be set to remove the poles of $\phi(z)$ that are inside the unit circle. This makes sure that $\phi(z)$ is analytic. The following lemma shows that the number of free constants that can be used in eliminating the inside poles is not the same as the total number of $\{\phi(\lambda_k)\}_{k=1}^{u}$ and $\{\phi^{(j)}(0)\}_{j=0}^{q}$, because of some of them may be linearly dependent on each other.

Lemma 3.3. *There exists a* $w \times N_1$ *matrix* $D_1(z)$*, a* $w \times 1$ *vector* $D_2(z)$*, and a* $N_1 \times 1$ *constant* *vector ψ, such that*

$$
\begin{aligned}
D\Big[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\Big] &= \widehat{D}_1(z)\Big[\phi(\lambda_1) \quad \ldots \quad \phi(\lambda_u) \quad \phi^0(0) \quad \ldots \quad \phi^q(0)\Big]' + D_2(z) \\
&= D_1(z)\psi + D_2(z) \qquad (3.23)
\end{aligned}
$$

where N_1 is the column rank of $\widehat{D}_1(z)$ and ψ is a linear combination of the constant vector

$$
[\phi(\lambda_1) \quad \ldots \quad \phi(\lambda_u) \quad \phi^0(0) \quad \ldots \quad \phi^q(0)]'.
$$

Proof. See Appendix A.8 for proof. □

Here, N_1 is the actual number of free constants that can used to remove the inside poles of $\phi(z)$. Theorem 3 shows that possible inside poles of $\phi(z)$ are from the inside roots of the determinant of $T(z)$. It follows that whether there exists a solution to model (3.8) or not hinges on whether there are enough free constants to eliminate all the inside roots of $\det[T(z)]$. Furthermore, there exists a unique solution if there are exactly N_1 conditions to determine the N_1 free constants.

Theorem 3 (General solution formula)**.** *Assume the signal process follows (3.1) and the model is given by equation (3.8). Let N_2 denote the number of roots of $\det[T(z)]$ that are inside the unit circle and let $\{\vartheta_1, \ldots, \vartheta_{N_2}\}$ denote these inside roots. Assume these roots are distinct. Define*

$$
U_1\psi + U_2 \equiv \begin{bmatrix} \det\Big[D_1(\vartheta_1)\psi + D_2(\vartheta_1) & T_1(\vartheta_1) & \ldots & T_{\ell_1-1}(\vartheta_1) & T_{\ell_1+1}(\vartheta_1) & \ldots & T_w(\vartheta_1)\Big] \\ \vdots & \vdots & \ldots & \vdots & \vdots & \ldots & \vdots \\ \det\Big[D_1(\vartheta_{N_2})\psi + D_2(\vartheta_{N_2}) & T_1(\vartheta_{N_{N_2}}) & \ldots & T_{\ell_{N_2}-1}(\vartheta_{N_2}) & T_{\ell_{N_2}+1}(\vartheta_{N_2}) & \ldots & T_w(\vartheta_{N_2})\Big] \end{bmatrix}
$$

where $T_{\ell_i}(\vartheta_i)$ is a linear combination of $\Big\{T_1(\vartheta_i), \ldots, T_{\ell_i-1}(\vartheta_i),\ T_{\ell_i+1}(\vartheta_i), \ldots, T_w(\vartheta_i)\Big\}$.

1. *If $N_1 < N_2$, there is no solution.*

2. *If $N_1 = N_2 = \text{rank}\ (U_1)$, there exists a unique solution $\phi(z)$. For $i \in \{1, \ldots, w\}$*

$$
\phi_i(z) = \frac{\det\Big[T_1(z) \ \ \ldots \ \ T_{i-1}(z) \ \ D_1(z)\psi + D_2(z) \ \ T_{i+1}(z) \ldots \ \ \ldots \ \ T_w(z)\Big]}{\det\Big[T(z)\Big]} \qquad (3.24)
$$

and

$$
\psi = -U_1^{-1}U_2 \qquad (3.25)
$$

3. *If* $N_1 > N_2$ *or* $N_1 = N_2 > rank\ (U_1)$*, there exists an infinite number of solutions.*

Proof. See Appendix A.9 for proof. □

With this theorem, we can prove our finite-state-representation theorem (Theorem 2), which is the last step of our method.

3.7 Innovation form and signal form

The solution we discussed in Section 3.6 is in terms of signals

$$y_t = h(L)x_t. \tag{3.26}$$

This is the most natural way to represent the policy rule because agents' actions depends on what they observe. However, sometimes it is more convenient to work with the policy function in terms of the underlying shocks.

$$y_t = d(L)s_t. \tag{3.27}$$

We label the solution in terms of signals as *signal form* and the solution in terms of underlying shocks as *innovation form*. Similar to the procedure to solve for $h(L)$, which effectively solves a system of equations in terms of signals, one can also write down the system of equations in terms of the underlying shocks $\{s_t\}$. A detailed description of the problem in innovation form can be found in Appendix A.11.

From a practical point of view, 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. However, the innovation form often provides a sharper characterization of the equilibrium, for the reason that the statistical properties are easier to derive in terms of the underlying shocks. Therefore, it is useful to obtain the solution in both forms. One may be concerned about whether the solution in innovation form is the same as the solution in signal form, and the following theorem shows that one can indeed work with either of them.

Theorem 4. *Assume the signal process follows (3.1) and the model is given by equation (3.8). There exists a solution in signal form,*

$$y_t = h(L)x_t, \tag{3.28}$$

if and only if there exists a solution in innovation form

$$y_t = d(L)s_t, \tag{3.29}$$

where $h(L)$ and $d(L)$ satisfy

$$d(L) = h(L)M(L)$$
$$h(L) = d(L)M'(L^{-1})\rho_{xx}(L)^{-1} - \sum_{k=1}^{u} \frac{d(\lambda_k)\lambda^k G(\lambda_k)V^{-1}B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k - \lambda_\tau)}$$

Proof. See Appendix A.12 for proof. □

If $M(L)$ 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, $\{y_t\}$ always lies in the space spanned by current and past signals because agents can only condition their choice on their observables, that is, , $\{y_t\} \subset \mathcal{H}_t^x \subset H_t^s$.

4 Application I: Two-Player Model

In this section, we use the method developed in Section 3 to solve two particular two-player games. There are only private signals in the first case, but we allow agents to share a common public signal in the second case.

4.1 Private Signal: Woodford (2003)

The model we use is akin to model (2.7) introduced in Section 3.

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

There is a continuum of agents, and each individual agent i's optimal choice satisfies equation (4.1). The aggregate action y_t is given by

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

We assume the economic fundamental ξ_t follows an AR(1) process

$$\xi_t = \rho\xi_{t-1} + \eta_t, \tag{4.3}$$

where $\eta_t \sim N(0,1)$ and we have normalized to the variance of η_t to be 1.

We assume that an agent i receives two private signals about ξ_t

$$x_{it}^1 = \xi_t + \epsilon_{it}, \tag{4.4}$$
$$x_{it}^2 = \xi_t + u_{it}, \tag{4.5}$$

where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $u_{it} \sim N(0, \sigma_u^2)$.

The equilibrium is defined as follows

Definition 4.1. *Given the signal process (4.3) to (4.5), the equilibrium of model (4.1) is a policy rule $h = \{h_1, h_2\} \in \ell^2 \times \ell^2$, such that*

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

where

$$y_{it} = h_1(L)x_{it}^1 + h_2(L)x_{it}^2,$$
$$y_t = \int y_{it}.$$

The structure of this model is similar to Woodford (2003). In Woodford (2003), y_{it} is the price chosen by an individual firm, y_t is the aggregate price level, and ξ_t can be interpreted as some aggregate demand shock. The focus of Woodford (2003) is that higher order beliefs generate inertia of the aggregate price level in response to the demand shock ξ_t (hump-shaped response), which is shown numerically. The following proposition gives the analytic solution to model (4.1), and the underlying reason for the inertia becomes transparent.

Proposition 4.1. *Assume that $\alpha \in (0, 1)$. Given the signal process (4.3) to (4.5), the equilibrium policy rule in model (4.1) is given by*

$$h_1(L) = \frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \tag{4.6}$$

$$h_2(L) = \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)}\frac{1}{1-\vartheta L}, \tag{4.7}$$

where

$$\vartheta = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)(\sigma_\epsilon^2+\sigma_u^2)}{\rho\sigma_\epsilon^2\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)(\sigma_\epsilon^2+\sigma_u^2)}{\rho\sigma_\epsilon^2\sigma_u^2}\right)^2 - 4}\,\right] \tag{4.8}$$

The finite-state representation is given by

$$y_{it} = \frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}x_{it}^1 + \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)}x_{it}^2 + z_{it}, \tag{4.9}$$

where

$$z_{it+1} = \vartheta z_{it} + \frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)} x_{it}^1 + \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)} x_{it}^2. \tag{4.10}$$

The aggregate y_t is given by

$$y_t = \frac{\vartheta}{\rho(1-\rho\vartheta)} \frac{\sigma_\epsilon^2 + \sigma_u^2}{\sigma_\epsilon^2 \sigma_u^2} \frac{1}{(1-\vartheta L)(1-\rho L)} \eta_t \tag{4.11}$$

Proof. See Appendix A.13 for proof. □

The individual policy rule follows an AR(1) process, and the aggregate y_t follows an AR(2) process. The two signals only differ by the variance of their idiosyncratic noises. As expected, $h_1(L)$ and $h_2(L)$ are symmetric, but the weight on each signal is adjusted according to their informativeness.

Crucially, the persistences of $h_1(L), h_2(L)$, and the persistence of aggregate y_t are governed by ϑ. Given ρ, as ϑ increases, the peak of the impulse response of y_t to η_t shifts to the right, which makes it possible to have a hump-shaped response. If ϑ is small enough, then there may not be any hump-shaped response. The following proposition provides a sharp characterization of ϑ.

Proposition 4.2. *Assume that $\alpha_1 \in (0,1), \rho \in (0,1), \sigma_\epsilon > 0$, and $\sigma_u > 0$. Then ϑ satisfies*

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

$$\lambda = \frac{1}{2}\left[\left(\frac{1}{\rho} + \rho + \frac{\sigma_\epsilon^2 + \sigma_u^2}{\rho\sigma_\epsilon^2\sigma_u^2}\right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{\sigma_\epsilon^2 + \sigma_u^2}{\rho\sigma_\epsilon^2\sigma_u^2}\right)^2 - 4}\right] \tag{4.12}$$

2. *ϑ is increasing in α and*

$$\lim_{\alpha_1 \to 1} \vartheta = \rho$$

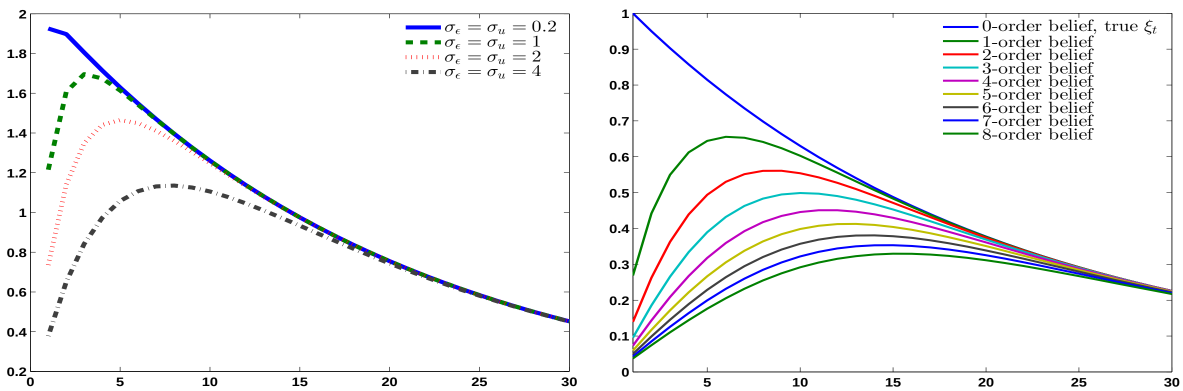
$$\lim_{\alpha_1 \to 0} \vartheta = \lambda$$

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, and with a large α, it is more likely for y_t to have a hump-shaped response . This is because with information frictions, higher order beliefs respond slowly to the shock. When the degree of strategic complementarity increases, higher order beliefs become more important in shaping the behavior of y_t, as shown in equation (2.10).

Example We use a numerical example to further illustrate the properties of the model economy. We set the degree of strategic complementarity $\alpha = 0.5$ and the persistence of ξ to be 0.95. As the variances of idiosyncratic shocks increase, the degree of information frictions also increases. As

FIGURE 2: Impulse Response to η Shock in the Private-Signal Model



Response of y_t with various σ_ϵ and σ_u Higher order bliefs with $\sigma_\epsilon = \sigma_u = 4$

shown in Figure 2, the hump-shaped response of y_t to η_t is more pronounced when there are larger information frictions. This is because ϑ is increasing in σ_ϵ and σ_u. When there is little information friction, ϑ is small and there is no hump-shaped response any more.

The higher order beliefs have the following feature: as the order increases, the higher order belief becomes less responsive, and the peak of its response shifts to the right. To predict ξ_t, agent i discounts his signals by the Kalman gain $1 - \lambda$, which leads to that $\mathbb{E}_{it}[\xi_t]$ is less volatility than ξ_t. When agent i infers others' forecasts of ξ_t, he realizes others also discount their signals by $1 - \lambda$. Agent i' best forecast of others signal is $\mathbb{E}_{it}[\xi_t]$, and his forecast of $\overline{\mathbb{E}}[\xi_t]$ in turn discounts the original ξ_t twice. This logic applies to all the higher order beliefs. Consider k-th order belief. As k increases, the forecasts of k-th order beliefs puts less weight on current signals, and more weight on the priors of beliefs with order lower than k, which makes the inertia increase in the order of beliefs.

4.2 Public Signal: Angeletos and La'O (2010)

Now we introduce the following variation to the model discussed in the last section. The economic fundamental ξ_t still follows an AR(1) process

$$\xi_t = \rho \xi_{t-1} + \eta_t, \tag{4.13}$$

but we assume the first signal about the economic fundamental ξ_t is the same across all the agents

$$x_{it}^1 = \xi_t + \epsilon_t, \tag{4.14}$$

$$x_{it}^2 = \xi_t + u_{it}, \tag{4.15}$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ and $u_{it} \sim N(0, \sigma_u^2)$. Note that now the noise in the first signal is not indexed by i. Effectively, the first signal now becomes a public signal. The model is the same as before

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

The structure of this model is similar to Angeletos and La'O (2010).[11] In Angeletos and La'O (2010), y_{it} is the output chosen by an individual firm i, y_t is the aggregate out, ξ_t is the aggregate TFP shock. The focus of their paper is to understand the effects of the common noise ϵ_t, which can be interpreted as animal spirits or sentiments. The question is whether this common noise can introduce aggregate output fluctuations. Angeletos and La'O (2010) use a guess-and-verify method and obtain a numerical solution. Here, we obtain an analytic solution.

Proposition 4.3. *Assume that $\alpha \in (0,1)$. Given the signal process (4.13) to (4.15), the equilibrium policy rule in model (4.16) is given by*

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

where

$$h_1(L) = \frac{1}{1-\alpha} \frac{\vartheta}{\rho \sigma_\epsilon^2 (1 - \rho\vartheta)} \frac{1}{1 - \vartheta L}, \tag{4.17}$$

$$h_2(L) = \frac{\vartheta}{\rho \sigma_u^2 (1 - \rho\vartheta)} \frac{1}{1 - \vartheta L}, \tag{4.18}$$

and

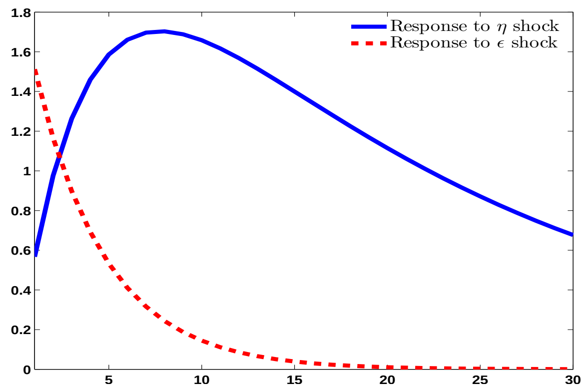
$$\vartheta = \frac{1}{2} \left[\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\epsilon^2 + \sigma_u^2}{\rho \sigma_\epsilon^2 \sigma_u^2} \right) - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{(1-\alpha)\sigma_\epsilon^2 + \sigma_u^2}{\rho \sigma_\epsilon^2 \sigma_u^2} \right)^2 - 4} \right] \tag{4.19}$$

The finite-state representation is given by

$$y_{it} = \frac{1}{1-\alpha} \frac{\vartheta}{\rho \sigma_\epsilon^2 (1 - \rho\vartheta)} x_{it}^1 + \frac{\vartheta}{\rho \sigma_u^2 (1 - \rho\vartheta)} x_{it}^2 + z_{it} \tag{4.20}$$

[11] The original model in Angeletos and La'O (2010) is $y_{it} = \xi_t + u_{it} + \alpha \mathbb{E}_{it}[y_t]$, where u_{it} is firm specific technology shock. Here, we modify their original model to better contrast with our private-signal model, but the main dynamics remain the same.

FIGURE 3: Impulse Response of y_t in the Public-Signal Model



where

$$z_{it+1} = \vartheta z_{it} + \frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}x_{it}^1 + \frac{\vartheta}{\rho\sigma_u^2(1-\rho\vartheta)}x_{it}^2 \tag{4.21}$$

The aggregate y_t *is given by*

$$y_t = \frac{\vartheta}{\rho(1-\rho\vartheta)}\frac{(1-\alpha)\sigma_\epsilon^2+\sigma_u^2}{(1-\alpha)\sigma_\epsilon^2\sigma_u^2}\frac{1}{(1-\vartheta L)(1-\rho L)}\eta_t + \frac{1}{1-\alpha}\frac{\vartheta}{\rho\sigma_\epsilon^2(1-\rho\vartheta)}\frac{1}{1-\vartheta L}\epsilon_t \tag{4.22}$$

Proof. See Appendix A.14 for proof. □

We can see that the public-model is clearly different from the private-signal model. Because the common noise ϵ_t in the first signal now affects all agents in the economy, each individual agent will respond more strongly to the first signal, due to the strategic complementarity. 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. In addition, σ_ϵ and σ_u are not symmetric in shaping the information frictions, reflected in how they affect the persistence ϑ in equation (4.19).

In terms of the aggregate y_t, it is now a function of both η shock and ϵ shock. However, the response to an η shock follows an AR(2) process, the same as the private-signal model, but the response to an ϵ shock follows an AR(2) process. Figure 3 plots the responses to these two shocks.[12]

[12]We set the degree of strategic complementarity α to be 0.5 and the persistence of ξ to be 0.95. We also set the variance of the noise to be $\sigma_\epsilon = \sigma_u = 4$. The implied persistence $\vartheta = 0.77$.

5 Application II: 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 (2013) 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 restrict their attention to the special case in which the number of signals equals the number of shocks and agents observe the endogenous variable without noise. 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

The model we use is similar to the private-signal model in Section 4.1, but we assume a different information structure. Agents still receive two signals. The first signal is the same as before, but the second one is the aggregate y_t with an idiosyncratic noise

$$x_{it}^2 = y_t + u_{it} = \int y_{jt} + u_{it} \tag{5.1}$$

The aggregate y_t is endogenously determined by all the individual choices, while at the same time, it is served as a signal for agents to infer the state of the economy. In this case, we find it is more convenient to define the equilibrium with innovation form.

Definition 5.1. *The equilibrium is an endogenous stochastic process Ω_{it}, a policy rule for an individual agent $\phi = \{\phi_1, \phi_2, \phi_3\} \in \ell^2 \times \ell^2 \times \ell^2$, and the law of motion for aggregate y_t, $\Phi \in \ell^2$, such that*

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

$$\begin{aligned} x_{it}^1 &= \xi_t + \epsilon_{it}, && (5.2) \\ x_{it}^2 &= y_t + u_{it}, && (5.3) \end{aligned}$$

where

$$\xi_t = \frac{\Pi_{k=1}^{n}(1+\kappa_k L)}{\Pi_{k=1}^{m}(1-\zeta_k L)}\eta_t, \tag{5.4}$$

$$y_t = \Phi(L)\eta_t. \tag{5.5}$$

2. *Individual rationality*

$$y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha\, \mathbb{E}_{it}[y_t], \tag{5.6}$$

where

$$y_{it} = \phi_1(L)\epsilon_{it} + \phi_2(L)u_{it} + \phi_3(L)\eta_t. \tag{5.7}$$

3. *Aggregate consistency:* $y_t = \int y_{it}$

$$\Phi(L) = \phi_3(L) \tag{5.8}$$

To show the generality of our claim, we allow ξ_t to follow any finite ARMA process. The equilibrium with endogenous information involves two fixed points. The first fixed point is individual rationality. Given the policy rule of others and the signal process, agent i optimally chooses the same policy rule as others. The second fixed point is absent in the equilibrium with exogenous information. 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 the cross-equation restriction in the sense that agents perception is in line with the reality generated by their own action.

Similar to Proposition 2.1, the following proposition guarantees that there exists a unique equilibrium with endogenous information.

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

Proof. See Appendix A.15 for proof. □

This proposition only proves the existence and uniqueness 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, even though there exists a unique equilibrium, the aggregate y_t does not follow a finite ARMA process. As a result, the equilibrium cannot have a finite-state representation.

Theorem 5. *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.16 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 exogenous information, Proposition 4.1 shows that if ξ_t follows an AR(1) process, the implied aggregate y_t follows an AR(2) process. With endogenous information, if we assume ξ_t follows an AR(1) process and the perceived y_t follows an AR(2) process, the implied actual y_t follows an ARMA $(4, 2)$ process. If we assume perceived y_t follows ARMA $(4, 2)$, the actual y_t will follow an ARMA $(6, 4)$ 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 (2013) 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 5 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 with an arbitrary ARMA process.

The reason for the infinite state variables, however, is not the infinite regress problem. 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 2: 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 (2013) 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 $x_t = M(L)s_t$ with $M(L)$ 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 is not complicated enough to create interesting dynamics. 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 arbitrarily well. The idea is to use a low order ARMA process to approximate aggregate y_t, which enables the Winer-Hopf prediction formula.

1. Assume that the perceived aggregate y_t follows an ARMA (p, q) process

$$y_t^p = \Phi(L)\eta_t = \sigma_y \frac{\Pi_{k=1}^q (1 + \theta_k L)}{\Pi_{k=1}^p (1 - \rho_k L)}.$$

2. Given the law of motion of the aggregate y_t, the signal process follows a finite ARMA process. Use the method in Section 3 to solve for the individual policy rule $\phi = \{\phi_1(L), \phi_2(L), \phi_3(L)\}$. The actual aggregate y_t follows

$$y_t^a = \phi_3(L)\eta_t.$$

3. To update $\Phi(L)$, expand $\phi_3(L)$ to obtain the infinite moving average representation. Choose the new $\sigma_y, \{\rho_k\}_{k=1}^p$ and $\{\theta_k\}_{k=1}^q$ to equate $\{\Phi_0, \Phi_1, \ldots, \Phi_{p+q}\}$ with $\{\phi_{30}, \phi_{31}, \ldots, \phi_{3p+q}\}$

4. Iterate 1 to 3 until the difference between $\{\Phi_0, \Phi_1, \ldots, \Phi_{p+q}\}$ and $\{\phi_{30}, \phi_{31}, \ldots, \phi_{3p+q}\}$ is smaller than the tolerance level.

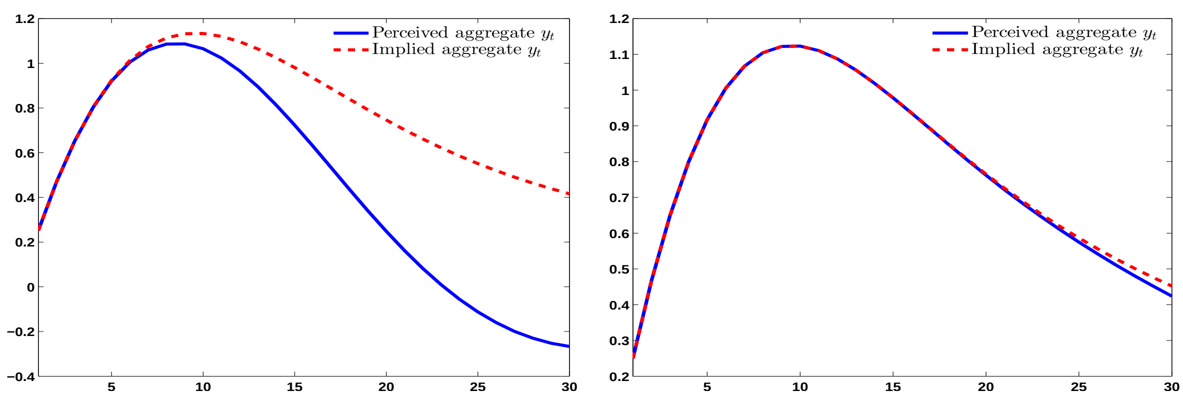
5. Compute $\|\Phi - \phi\|$ (one can simply use the norm of ℓ^2). If $\|\Phi - \phi\|$ is larger than the tolerance level, increase (p, q) and repeat 1 to 4.

Based on the proof of Proposition A.15, this algorithm is a contraction mapping that converges to the true solution as the order of the ARMA approximation increases. 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 model, 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 (2011), our method has the following advantage. The first advantage is that our method requires fewer state variables. Nirmark's method needs to keep track of a large number of higher order beliefs to accurately approximate the policy rule. In principle, Nirmark's method is to use MA(∞) process to approximate the policy rule while our method uses an ARMA process for approximation, which is more efficient. 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 easier to implement and is applicable in more general environments. Nirmark's method relies on the correct conjecture of the law of motion of the higher order beliefs. When the signal process is more complicated than an AR(1) process, it is not obvious what the correct conjecture should be. In addition, Nirmark's method also relies on that the equilibrium policy rule is a relatively simple function of higher order beliefs, but this may not be true in many empirical applications where the system is complicated (see the quantitative model in Huo and Takayama (2014) for example). Instead, our method instead does not hinge on these assumptions.

Example To check whether our approximation method is accurate enough, we need to compare the perceived aggregate y_t with the implied aggregate y_t. We set $\alpha = 0.5$. We assume ξ_t follows an AR(1) process with persistence $\rho = 0.95$. We also set $\sigma_a = \sigma_u = 4$. As shown in Figure 4, if we use an AR(2) process to approximate the aggregate y_t, the difference between perceived and implied aggregate y_t is quite noticeable. If we use an ARMA (4,2) process to approximate y_t, the perceived and implied y_t are almost identical to each other. 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.

FIGURE 4: Comparing Approximation Accuracy for $\alpha = 0.8$



AR(2) approximation for y_t ARMA(4,2) approximation for y_t

6 Application III: Random-Matching Model, Angeletos and La'O (2013)

In this section we discuss another type of model, in which an agent meets a different player every period. Angeletos and La'O (2013) consider an interesting model environment with this feature, but they assume there is no persistent shock in their baseline model. 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, and it makes the model unsuitable for empirical work. We extend Angeletos and La'O (2013) to allow persistent shocks and the infinite regress problem in the model.

Assume that there is a continuum of agents in the economy. An individual agent i is endowed with a productivity a_i, which is drawn from a normal distribution $N(0, \sigma_a^2)$. Note that both individual's productivity and the distribution is fixed over time, and there is no aggregate uncertainty with respect to the economic fundamentals. At the beginning of each period, an agent i is randomly matched with another agent $m(i,t)$ and trades goods with $m(i,t)$, where $m(i,t)$ is the index of agent i's trading partner in period t. Note that the production has to take place before trading, and agents have to infer others' output based on their signals. Due to strategic complementarity, agent i's optimal output choice y_{it} needs to satisfy

$$y_{it} = a_i + \alpha\, \mathbb{E}_{it}[y_{m(i,t)t}], \tag{6.1}$$

where $\alpha \in (0,1)$ controls the degree of strategic complementarity, and $y_{m(i,t)t}$ is the output choice of i's trading partner $m(i,t)$ at period t. Equation (6.1) says that agent i' output is increasing in his own productivity and the output of his trading partner, while the detailed micro-foundation that leads to this equation is not important for us to understand the infinite regress problem.

Agent i receives two signals

$$x_{it}^1 = a_{m(i,t)} + \epsilon_{it}, \tag{6.2}$$

$$x_{it}^2 = x_{m(i,t)t}^1 + \xi_t + u_{it}, \tag{6.3}$$

where $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$ and $u_{it} \sim N(0, \sigma_u^2)$, both of which are idiosyncratic noise. The productivity of i's trading partner is $a_{m(i,t)}$, and from i's perspective, it is also an i.i.d shock that follows $N(0, \sigma_a^2)$. $x_{m(i,t)t}^1$ is the first signal received by agent i's trading partner $m(i,t)$. ξ_t is common for all agents, which follows an AR(1) process

$$\xi_t = \rho \xi_{t-1} + \eta_t, \tag{6.4}$$

where $\eta_t \sim N(0,1)$. In Angeletos and La'O (2013), ξ_t is an i.i.d shock, but we assume $\rho \in (0,1)$ here to introduce the infinite regress problem. The information set of agent i is[13]

$$\Omega_{it} = \left\{ a_i,\ x_{it}^1,\ x_{it}^2,\ x_{it-1}^1,\ x_{it-1}^2,\ x_{it-2}^1,\ x_{it-2}^2,\ \dots \right\}. \tag{6.5}$$

Note that a_i needs to be included in the information set because agent i's action directly depends on a_i, and it also helps to predict i's trading partner's signal. The equilibrium is defined as

Definition 6.1. *Given the signal process (6.2) to (6.4), the equilibrium of model (6.1) is a policy rule $h = \{h_a, h_1, h_2\} \in \mathbb{R} \times \ell^2 \times \ell^2$, such that*

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

where

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

As emphasized by Angeletos and La'O (2013), agent i's estimate of his trading partners' productivity $a_{m(i,t)}$ is pinned downed by the i's first signal alone, and not affected by the second signal. However, agent i's estimate of $x_{m(i,t)t}^1$ is affected by the common noise ξ_t. With a positive realization of ξ_t, agent i attributes part of ξ_t to $x_{m(i,t)t}^1$, and believes that agent $m(i,t)$ will overestimate i's productivity a_i and produce more output. Therefore, agent i's also optimally produces more output due to strategic complementarity. In aggregate, ξ_t leads to fluctuations in total output by affecting all agents' higher order beliefs.

Different from the applications discussed in Section 4, agent i has to form higher order beliefs about a random player $m(i,t)$ every period. This change may prevent the use of the guess-and-verify method, but our method developed in Section 3 can still be applied to solve the model.

Proposition 6.1. *Assume that $\alpha \in (0,1)$. Given the signal process (6.2) to (6.4), the equilibrium policy rule in model (6.1) is given by*

$$h_a = 1 + \alpha\varphi - \frac{\alpha\vartheta\varphi(1-\rho)}{\rho(1-\vartheta)} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2} \tag{6.6}$$

$$h_1(L) = \varphi \tag{6.7}$$

$$h_2(L) = \frac{\alpha\vartheta\varphi}{\rho} \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2} \frac{1-\rho L}{1-\vartheta L} \tag{6.8}$$

[13] There is an implicit assumption that agents do not observe their trading partner's output or productivity level after production. This assumption is only to implement the notion of imperfect communication between producers and transactors, but is not important for our purpose.

where

$$\vartheta = \frac{1}{2}\left[\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho(\sigma_\epsilon^2 + \sigma_u^2)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho(\sigma_\epsilon^2 + \sigma_u^2)}\right)^2 - 4}\right], \tag{6.9}$$

$$\varphi = \frac{\alpha}{1 - \alpha^2 + \frac{\sigma_\epsilon^2}{\sigma_a^2}\left(1 - \alpha^2 \frac{\vartheta}{\rho}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2}\right)}. \tag{6.10}$$

The finite-state representation is given by

$$y_{it} = h_a a_i + \varphi x_{it}^1 + \frac{\alpha\vartheta\varphi}{\rho}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2} x_{it}^2 + z_{it} \tag{6.11}$$

where

$$z_{it+1} = \vartheta z_{it} + \frac{(1-\rho)\alpha\vartheta\varphi}{\rho}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2} x_{it}^2 \tag{6.12}$$

The aggregate y_t is given by

$$y_t = \frac{\alpha\vartheta\varphi}{\rho}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_u^2}\frac{1}{1-\vartheta L}\eta_t \tag{6.13}$$

Proof. See Appendix A.17 for proof. □

Note that $h_1(L)$ is a constant, which implies that the policy rule does not depend on $\{x_{i\tau}^1\}_{\tau=-\infty}^{t-1}$. The reason is that the first signal is only useful in predicting the productivity of current trading partner, but it is independent of the persistent shock ξ_t. It turns out that $h_2(L)$ follows an ARMA(1,1) process, and the aggregate output y_t follows an AR(1) process.

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. Assume that agent i observes both ξ_t and $a_{m(i,t)t}$ perfectly. However, agent i believes his trading partner $m(i,t)$ observes a_i with bias ξ_t. If 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 output of his trading partner is

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

In equilibrium,

$$y_{it} = \alpha_0 a_i + \alpha_1 \mathbb{E}_{it}[y_{m(i,t)t}],$$

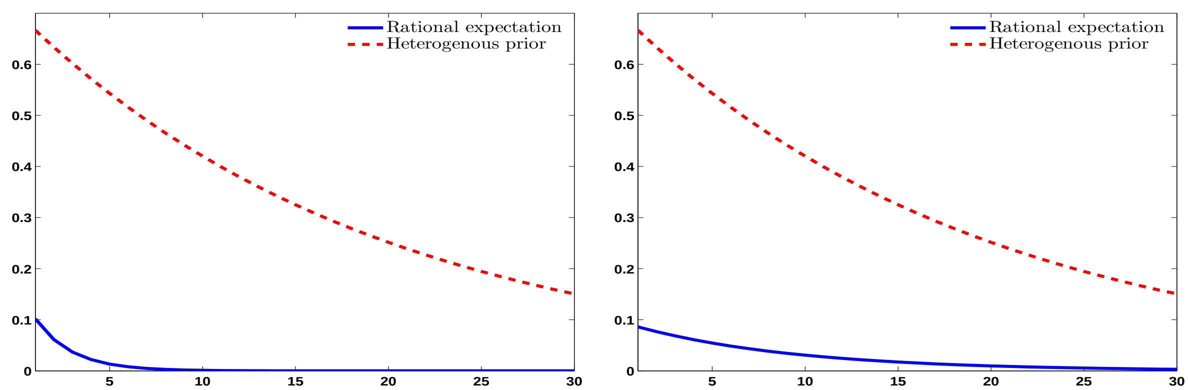
which leads to

$$y_{it} = \frac{1}{1-\alpha^2} a_i + \frac{\alpha}{1-\alpha^2} a_{m(i,t)} + \frac{\alpha_1^2}{(1-\alpha_1^2)(1-\alpha)} \xi_t \tag{6.14}$$

$$y_t = \frac{\alpha^2}{(1-\alpha^2)(1-\alpha)} \xi_t \tag{6.15}$$

Quantitatively, by assuming heterogeneous prior, y_t is perfectly correlated with ξ_t, while in our model with common prior, the persistence of aggregate output 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 5.

FIGURE 5: Impulse Response of y_t to η Shock in Random-Player Model



Samll information friction. $\sigma_\epsilon = \sigma_u = 1$ Large information friction. $\sigma_\epsilon = \sigma_u = 4$

Note that the both the persistence and instantaneous response of y_t under heterogeneous prior is very different from the solution under rational expectation. The solution under heterogeneous prior assumption is independent of the degree of information frictions, that is, the distribution of idiosyncratic productivity and the size of the idiosyncratic noise do not affect the behaviour of output. By assuming heterogeneous prior, one effectively assumes away the information frictions, which is the reason that higher order beliefs arise in the first place. The method we provide to solve the infinite regress problem retains the notion of rationality, and we can pin down the degree of information frictions by comparing the model results with data.

7 Application IV: a Quantitative Business Cycle Model

Application I to Application III can be thought of as various extensions of the basic model presented in Section 2. These applications are theoretically interesting, but have not fully taken advantage of our method. In the general model structure we outline in equation (3.8), we allow the model to include the past, the present, and the future values of the choice variables, the choices of others, and the exogenous variables.

In a companion paper (Huo and Takayama, 2014), we apply our method and solve a full-blown quantitative model in which the confidence shock alone is sufficient to account for the main aggregates in business cycles. The idea is related to Angeletos and La'O (2013), but our model differs from theirs in several crucial ways. We maintain the strong notion of rationality and solve the infinite regress problem directly. Agents need to choose both labor and investment, and need to infer the output and capital of both their current and future trading partners. There are multiple persistent shocks in the signal process to match various micro and macro moments. Therefore, higher order beliefs affect agents' decisions in a fairly complex way. With our preferred calibration of information frictions, we find that the model with confidence shocks generate much of the volatility and co-movement of aggregate variables, but it has difficulty in matching the persistence of the aggregate variables.

8 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. Preliminary findings in Huo and Takayama (2014) show that this is a promising direction to pursue.

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.

Hansen, L. P., and T. J. Sargent (1980): "Linear rational expectations models for dynamically interrelated variables," Discussion paper.

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

Huo, Z., and N. Takayama (2014): "Higher Order Beliefs, Confidence and Business Cycles," Manuscript, University of Minnesota.

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.

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

——— (2011): "Dynamic higher order expectations," Economics Working Papers 1118, Department of Economics and Business, Universitat Pompeu Fabra.

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

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

Rondina, G., and T. B. Walker (2013): "Information Equilibria in Dynamic Economies," 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 Proof of Proposition 2.1

Proof. We consider the equilibrium in the innovation form. Let $\phi = \{\phi_1, \phi_2, \phi_3\} \in \ell^2 \times \ell^2 \times \ell^2$. The norm of ϕ can de defined as

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

Given ϕ, let

$$y_{it} = \phi_1(L)\epsilon_{it} + \phi_2(L)u_{it} + \phi_3(L)\eta_t,$$

and let

$$\mathbb{E}_{it}[\phi_1(L)\epsilon_{jt} + \phi_2(L)u_{jt} + \phi_3(L)\eta_t] \equiv \widehat{\phi}_1(L)\epsilon_{it} + \widehat{\phi}_2(L)u_{it} + \widehat{\phi}_3(L)\eta_t$$

The inference of ξ_t is independent of ϕ and is given by

$$\mathbb{E}_{it}[\xi_t] \equiv g_1(L)\epsilon_{it} + g_2(L)u_{it} + g_3(L)\eta_t.$$

If $y_{it} = g_1(L)\epsilon_{it} + g_2(L)u_{it} + g_3(L)\eta_t + \alpha\left(\widehat{\phi}_1(L)\epsilon_{it} + \widehat{\phi}_2(L)u_{it} + \widehat{\phi}_3(L)\eta_t\right)$, then ϕ is an equilibrium.

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

$$\mathcal{T}(\phi) = \mathcal{T}(\{\phi_1, \phi_2, \phi_3\}) = \{g_1 + \alpha\widehat{\phi}_1, g_2 + \alpha\widehat{\phi}_2, g_3 + \alpha\widehat{\phi}_3\}$$

The equilibrium is a fixed point of the operator $\mathcal{T}$. If we can show that $\mathcal{T}$ is a contraction mapping, it is sufficient to prove the theorem.

Let $\phi \in \ell^2 \times \ell^2 \times \ell^2$ and $\psi \in \ell^2 \times \ell^2 \times \ell^2$. The distance between ϕ and ψ is

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

The distance between $\mathcal{T}(\phi)$ and $\mathcal{T}(\psi)$ is

$$\|\mathcal{T}(\phi) - \mathcal{T}(\psi)\| = |\alpha| \sqrt{\sigma_\epsilon^2 \sum_{i=0}^{\infty} (\widehat{\phi}_{1i} - \widehat{\psi}_{1i})^2 + \sigma_u^2 \sum_{i=0}^{\infty} (\widehat{\phi}_{2i} - \widehat{\psi}_{2i})^2 + \sigma_\eta^2 \sum_{i=0}^{\infty} (\widehat{\phi}_{3i} - \widehat{\psi}_{3i})^2}$$

Note that the variance of a variable is always larger than the variance of its predictor

$$\begin{aligned} &\mathrm{Var}\Big[[\phi_1(L) - \psi_1(L)]\epsilon_{jt} + [\phi_2(L) - \psi_2(L)]u_{jt} + [\phi_3(L) - \psi_3(L)]\eta_t\Big] \\ \geq &\mathrm{Var}\Big[\mathbb{E}_{it}\Big[[\phi_1(L) - \psi_1(L)]\epsilon_{jt} + [\phi_2(L) - \psi_2(L)]u_{jt} + [\phi_3(L) - \psi_3(L)]\eta_t\Big]\Big] \end{aligned}$$

We have

$$
\begin{aligned}
&\text{Var}\Big[[\phi_1(L)-\psi_1(L)]\epsilon_{jt}+[\phi_2(L)-\psi_2(L)]u_{jt}+[\phi_3(L)-\psi_3(L)]\eta_t\Big]\\
=&\sigma_\epsilon^2\sum_{k=0}^{\infty}(\phi_{1k}-\psi_{1k})^2+\sigma_u^2\sum_{k=0}^{\infty}(\phi_{2k}-\psi_{2k})^2+\sigma_\eta^2\sum_{k=0}^{\infty}(\phi_{3k}-\psi_{3k})^2\\
=&\|\phi-\psi\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\text{Var}\Big[\mathbb{E}_{it}\Big[[\phi_1(L)-\psi_1(L)]\epsilon_{jt}+[\phi_2(L)-\psi_2(L)]u_{jt}+[\phi_3(L)-\psi_3(L)]\eta_t\Big]\Big]\\
=&\text{Var}\Big[[\widehat{\phi}_1(L)-\widehat{\psi}_1(L)]\epsilon_{it}+[\widehat{\phi}_2(L)-\widehat{\psi}_2(L)]u_{it}+[\widehat{\phi}_3(L)-\widehat{\psi}_3(L)]\eta_t\Big]\\
=&\sigma_\epsilon^2\sum_{i=0}^{\infty}(\widehat{\phi}_{1i}-\widehat{\psi}_{1i})^2+\sigma_u^2\sum_{i=0}^{\infty}(\widehat{\phi}_{2i}-\widehat{\psi}_{2i})^2+\sigma_\eta^2\sum_{i=0}^{\infty}(\widehat{\phi}_{3i}-\widehat{\psi}_{3i})^2\\
=&\|\mathcal{T}(\phi)-\mathcal{T}(\psi)\|^2\left(\frac{1}{\alpha}\right)^2.
\end{aligned}
$$

Therefore, $\|\mathcal{T}(\phi)-\mathcal{T}(\psi)\|\leq\alpha\|\phi-\psi\|$. When $\alpha\in(0,1)$, the operator $\mathcal{T}$ is a contraction mapping, and there exists a unique fixed point. □

A.2 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.3 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.

Let $r_{ij} = \max\{p_{ij}, q_{ij} + 1\}$, and let $\alpha_{ijk} = 0$ if $k > q_{ij}$ and $\beta_{ijk} = 0$ if $k > q_{ij}$. Let $r = \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij}$. F is a $r \times r$ matrix with the following form

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

The element F_{ij} in F is a $r_{ij} \times r_{ij}$ matrix

$$F_{ij} = \begin{bmatrix} \alpha_{ij1} & \alpha_{ij2} & \dots & \alpha_{ijr_{ij}-1} & \alpha_{ijr_{ij}} \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Q is a $r \times m$ matrix with the following form

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

The element D_{ij} in D is a $r_{ij} \times m$ matrix

$$Q_{ij} = \begin{bmatrix} 0 & \dots & \alpha_{ij0} & \dots & 0 \\ 0 & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}, \tag{A.3}$$

where α_{ij0} is at the jth column.

H is a $n \times r$ matrix with the following form

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

The element H_{ij} in H is a $1 \times r_{ij}$ matrix

$$H_{ij} = \begin{bmatrix} 1 & \beta_{ij1} & \beta_{ij2} & \dots & \beta_{ijr_{ij}} \end{bmatrix}.$$

Let Z_t follows

$$Z_t = FZ_{t-1} + Qs_t.$$

We have

$$x_t = M(L)s_t = HZ_t$$

To show that the eigenvalues of F lie inside the unit circle, we can iterate the Z_t to obtain

$$Z_t = \sum_{j=0}^{\infty} F^j L^j Q s_t = (I - FL)^{-1} Q s_t$$

If the eigenvalues of F lies outside the unit circle, it follows that Z_t is not co-variance stationary, which contradicts the assumption that x_t is co-variance stationary. □

A.4 Proof of Theorem 3.5

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 x^∞ instead of x^t. The optimal linear prediction of y_t is

$$E[y_t|x^\infty] = \rho_{yx}(L)\rho_{xx}(L)^{-1}x_t.$$

This formula resembles the familiar formula in OLS regression. ρ_{yx} measures the correlation between y and x, adjusted by ρ_{xx}. Given the fundamental representation

$$x_t = B(L)w_t,$$

the prediction is equivalent to the prediction conditional on w^∞ and the prediction formula can be written as

$$\begin{aligned}
E[y_t|x^\infty] =& E[y_t|w^\infty] \\
=& \rho_{yx}(L)\rho_{xx}(L)^{-1}x_t, \\
=& \rho_{yx}(L)B'(L^{-1})^{-1}V^{-1}B(L)^{-1}B(L)w_t, \\
=& \rho_{yx}(L)B'(L^{-1})^{-1}V^{-1}w_t.
\end{aligned}$$

Now imagine the prediction is conditional on only current and past signals x^t, which is equivalent to conditional on w^t. Since w_t is serially uncorrelated, the best forecast of w_i for $i > t$ is zero. Note that $\rho_{yx}(L)B'(L^{-1})^{-1}$ contains negative powers of L and the best forecast of w_i for $i > t$ is zero, the optimal prediction for y_t is simply

$$\begin{aligned}
E[y_t|x^t] =& E[y_t|w^t] \\
=& [\rho_{yx}(L)B'(L^{-1})^{-1}]_+V^{-1}w_t, \\
=& [\rho_{yx}(L)B'(L^{-1})^{-1}]_+V^{-1}B(L)^{-1}x_t, \\
=& [\rho_{yx}(L)B'(L^{-1})^{-1}]_+V^{-1}B(L)^{-1}M(L)s_t.
\end{aligned}$$

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

A.5 Proof of Lemma 3.2

Proof. By the Canonical Factorization Theorem, it follows that the inverse of $B(z)$ is given by

$$\begin{aligned}
B(z)^{-1} =& I_n - H[I_r - (F - FKH)z]^{-1}FKz \\
=& \frac{I_n \det[I_r - (F - FKH)z] - H\mathrm{Adj}[I_r - (F - FKH)z]FKz}{\det[I_r - (F - FKH)z]} \\
=& \frac{\widehat{B}(z)}{\Pi_{k=1}^u (1 - \lambda_k z)}
\end{aligned} \tag{A.5}$$

where $\widehat{B}(z)$ is a matrix and the elements are all polynomials in z with finite degree, u is the degree of $\det[I_r - (F - FKH)z]$, and $\{\lambda_k\}_{k=1}^u$ are non-zero eigenvalues of $F - FKH$. To see why this is true, note that if λ_k is the eigenvalue of $F - FKH$, it satisfies

$$\det[\lambda_k I_r - (F - FKH)] = 0$$

which implies

$$\det\left[I_r - (F - FKH)\frac{1}{\lambda_k}\right] = 0$$

That is, $\frac{1}{\lambda_i}$ is the root of the determinant of $I_r - (F - FKH)z$. Reversely, the roots of $I_r - (F - FKH)z$ must be the reciprocals of the non-zero eigenvalues of $F - FKH$. In addition, Theorem 3.4 guarantees all of these eigenvalues of $F - FKH$ lie inside the unit circle.

Meanwhile, we have

$$B(z) = I_n + H[I_r - Fz]^{-1}FKz,$$

and

$$\begin{aligned}
B(z)^{-1} =& \left[I_n + H[I_r - Fz]^{-1}FKz\right]^{-1} \\
=& \left[\frac{I_n \det[I_r - Fz] - H\text{Adj}[I_r - Fz]FKz}{\det[I_r - Fz]}\right]^{-1} \\
=& \det[I_r - Fz]\left[I_n \det[I_r - Fz] - H\text{Adj}[I_r - Fz]FKz\right]^{-1}
\end{aligned}$$

Note that equation (A.5) has to be satisfied at the same time. As a result, there exists a matrix $\widehat{C}(z)$ such that the elements of it are all finite polynomials in z, and

$$B(z)^{-1} = \det[I_r - Fz]\frac{\widehat{C}(z)}{\Pi_{k=1}^{u}(1 - \lambda_k z)}. \tag{A.6}$$

The roots of $\det[I_r - Fz]$, which are the inverse of the eigenvalues of F, are different from $\{\lambda_k\}_{k=1}^{u}$, which are the inverse of the eigenvalues of $F - FKH$. By construction, the degree of $\Pi_{k=1}^{u}(1 - \lambda_k z)$ is larger than the degree of $\det[I_r - Fz]\widehat{C}(z)$.

By Lemma 3.1,

$$x_t = M(L)s_t = HZ_t = H(I_r - FL)^{-1}s_t \tag{A.7}$$

Combining equation (A.6) and (A.7) leads to

$$\begin{aligned}
B(z)^{-1}M(z) =& \det[I_r - Fz]\frac{\widehat{C}(z)}{\Pi_{k=1}^{u}(1 - \lambda_k z)}\ H(I_r - Fz)^{-1} \\
=& \det[I_r - Fz]\frac{\widehat{C}(z)}{\Pi_{k=1}^{u}(1 - \lambda_k z)}\ H\ \frac{\text{Adj}[I_r - Fz]}{\det[I_r - Fz]} \\
=& \frac{\widehat{C}(z)H\text{Adj}[I_r - Fz]}{\Pi_{k=1}^{u}(1 - \lambda_k z)}.
\end{aligned}$$

Here, the numerator of $B(z)^{-1}M(z)$ is a finite polynomial in z, and the degree of $\det[I_r - Fz]$ is larger than the degree of

$\text{Adj}[I_r - Fz]$. Therefore,

$$
\begin{aligned}
M'(z^{-1})B'(z^{-1})^{-1} =& \frac{\left(\widehat{C}(z^{-1})H\text{Adj}[I_r - Fz^{-1}]\right)'}{\Pi_{k=1}^{u}(1-\lambda_k z^{-1})} = \frac{\left(z^u\widehat{C}(z^{-1})H\text{Adj}[I_r - Fz^{-1}]\right)'}{\Pi_{k=1}^{u}(z-\lambda_k)} \\
=& \frac{G(z)}{\Pi_{k=1}^{u}(z-\lambda_k z)},
\end{aligned}
$$

where $G(z)$ is a polynomial in z because the degree of $\widehat{C}(z)H\text{Adj}[I_r - Fz]$ is less than u. □

A.6 Proof of Proposition 3.1

Proof. By the Wiener-Hopf Theorem, the prediction formula is

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

We need to obtain the formula for

$$
\left[\psi(L)M'(L^{-1})B'(L^{-1})^{-1}\right]_+ = \sum_{i=1}^{m}\begin{bmatrix} \frac{1}{\Pi_{k=1}^{u}(L-\lambda_k)}\psi_i(L)G_{i1}(L) \\ \vdots \\ \frac{1}{\Pi_{k=1}^{u}(L-\lambda_k)}\psi_i(L)G_{in}(L)\end{bmatrix}_+' \tag{A.8}
$$

Suppose $g(z)$ is a rational function of z that does not contains negative powers of z in expansion, then

$$
\left[\frac{g(z)}{(z-\lambda_1)\cdots(z-\lambda_u)}\right]_+ = \frac{g(z)}{(z-\lambda_1)\cdots(z-\lambda_u)} - \sum_{k=1}^{u}\frac{g(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}
$$

It follows that

$$
\begin{aligned}
\left[\psi(L)M'(L^{-1})B'(L^{-1})^{-1}\right]_+ &= \sum_{i=1}^{m}\begin{bmatrix} \frac{1}{\Pi_{k=1}^{u}(L-\lambda_k)}\psi_i(L)G_{i1}(L) - \sum_{k=1}^{u}\frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} \\ \vdots \\ \frac{1}{\Pi_{k=1}^{u}(L-\lambda_k)}\psi_i(L)G_{in}(L) - \sum_{k=1}^{u}\frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\end{bmatrix}' \\
= \quad \psi(L)M'(L^{-1})B'(L^{-1})^{-1} &- \sum_{i=1}^{m}\begin{bmatrix} \sum_{k=1}^{u}\frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} \\ \vdots \\ \sum_{k=1}^{u}\frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\end{bmatrix}'
\end{aligned}
$$

Also note that if $g(z) = [f(z)]_+$, then for $j = \{1, 2, \ldots\}$

$$
\begin{aligned}
&[z^{-j} f(z)]_+ \\
=&\left[z^{-j}[g(z) + f(z) - g(z)]\right]_+ \\
=&[z^{-j} g(z)]_+ + \left[z^{-j}[f(z) - g(z)]\right]_+ \\
=&z^{-j}\left(g(z) - \sum_{p=0}^{j-1} p!\, z^p [g(z)]_0^{(p)}\right)
\end{aligned}
$$

where $[g(z)]_0^{(p)}$ denotes p-th order derivative evaluated at 0. Applying this formula, we have,

$$
\begin{aligned}
&[L^{-j}\psi(L)G(L)]_+ \\
= & \sum_{i=1}^{m}\begin{bmatrix}
\left(\frac{\psi_i(L)G_{i1}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} - \sum_{p=0}^{j-1} p!\, L^p \left[\frac{\psi_i(L)G_{i1}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(p)}\right)L^{-j} \\
\vdots \\
\left(\frac{\psi_i(L)G_{in}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} - \sum_{p=0}^{j-1} p!\, L^p \left[\frac{\psi_i(L)G_{in}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(p)}\right)L^{-j}
\end{bmatrix}' \\
= & \psi(L)L^{-j}M'(L^{-1})B'(L^{-1})^{-1} \\
& - \begin{bmatrix}
\left(\sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} + \sum_{p=0}^{j-1} p!\, L^p \left[\frac{\psi_i(L)G_{i1}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{i1}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(p)}\right)L^{-j} \\
\vdots \\
\left(\sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} + \sum_{p=0}^{j-1} p!\, L^p \left[\frac{\psi_i(L)G_{in}(L)}{\Pi_{k=1}^u(L-\lambda_k)} - \sum_{k=1}^u \frac{\psi_i(\lambda_k)G_{in}(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(p)}\right)L^{-j}
\end{bmatrix}'
\end{aligned}
$$

□

A.7 Proof of Proposition 3.2

Proof. By Proposition 3.1, the system (3.8) can be written as

$$\begin{bmatrix} \phi(L)\left(\sum_{j=-q}^{p}\sum_{i=1}^{r} C_{1,i}^{y,j} A_i L^j\right) x_t \\ \vdots \\ \phi(L)\left(\sum_{j=-q}^{p}\sum_{i=1}^{r} C_{r,i}^{y,j} A_i L^j\right) x_t \end{bmatrix} + \begin{bmatrix} \phi(L)\left(\sum_{j=-q}^{p}\sum_{i=1}^{v} C_{1,i}^{f,j} f_i(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1}\right) x_t \\ \vdots \\ \phi(L)\left(\sum_{j=-q}^{p}\sum_{i=1}^{v} C_{r,i}^{f,j} f_i(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1}\right) x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{j=-q}^{p} C_1^{g,j} g(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1} x_t \\ \vdots \\ \sum_{j=-q}^{p} C_r^{g,j} g(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1} x_t \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{u} \frac{\phi(\lambda_k)\left(\sum_{j=-q}^{p}\sum_{i=1}^{v} \lambda_k^j C_{1,i}^{f,j} f_i(\lambda_k) G(\lambda_k) V^{-1} B(L)^{-1}\right)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \\ \vdots \\ \sum_{k=1}^{u} \frac{\phi(\lambda_k)\left(\sum_{j=-q}^{p}\sum_{i=1}^{v} \lambda_k^j C_{r,i}^{f,j} f_i(\lambda_k) G(\lambda_k) V^{-1} B(L)^{-1}\right)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \end{bmatrix} + \begin{bmatrix} \sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j C_1^{g,j} g(\lambda_k) G(\lambda_k) V^{-1} B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \\ \vdots \\ \sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j C_r^{g,j} g(\lambda_k) G(\lambda_k) V^{-1} B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{\sum_{i=1}^{r} \phi(L) C_{1,i}^{y,-j} A_i M(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r} \phi(\lambda_k) C_{1,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \\ \vdots \\ \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{\sum_{i=1}^{r} \phi(L) C_{r,i}^{y,-j} A_i M(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r} \phi(\lambda_k) C_{r,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{\sum_{i=1}^{v} \phi(L) C_{1,i}^{f,-j} f_i(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r} \phi(\lambda_k) C_{1,i}^{y,-j} f_i(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \\ \vdots \\ \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{\sum_{i=1}^{v} \phi(L) C_{r,i}^{f,-j} f_i(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r} \phi(\lambda_k) C_{r,i}^{f,-j} f_i(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{C_{1,i}^{g,-j} g(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{1,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \\ \vdots \\ \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \left[\frac{C_{r,i}^{g,-j} g(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{r,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} V^{-1} B(L)^{-1} x_t \end{bmatrix}$$

Rearranging the system of equations above to isolate $\phi(L)$ leads to the following more compact way

$$\begin{bmatrix} \phi(L)\sum_{j=-q}^{p} L^j \left[\sum_{i=1}^{r} C_{1,i}^{y,j} A_i + \sum_{i=1}^{v} C_{1,i}^{f,j} f_i(L) M'(L^{-1})\rho_{xx}(L)^{-1}\right] x_t \\ \vdots \\ \phi(L)\sum_{j=-q}^{p} L^j \left[\sum_{i=1}^{r} C_{r,i}^{y,j} A_i + \sum_{i=1}^{v} C_{r,i}^{f,j} f_i(L) M'(L^{-1})\rho_{xx}(L)^{-1}\right] x_t \end{bmatrix}$$

$$= -\begin{bmatrix} \sum_{j=-q}^{p} C_1^{g,j} g(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1} x_t \\ \vdots \\ \sum_{j=-q}^{p} C_r^{g,j} g(L) L^j M'(L^{-1})\rho_{xx}(L)^{-1} x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j \left[\sum_{i=1}^{v} C_{1,i}^{f,j}\phi(\lambda_k) f_i(\lambda_k) + \sum_{i=1}^{v2} C_1^{g,j} g(\lambda_k)\right] G(\lambda_k) V^{-1} B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \\ \vdots \\ \sum_{k=1}^{u} \frac{\sum_{j=-q}^{p} \lambda_k^j \left[\sum_{i=1}^{v} C_{r,i}^{f,j}\phi(\lambda_k) f_i(\lambda_k) + \sum_{i=1}^{v2} C_r^{g,j} g(\lambda_k)\right] G(\lambda_k) V^{-1} B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)} x_t \end{bmatrix}$$

$$+ \begin{bmatrix} \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \Bigg(\left[\frac{\sum_{i=1}^{r}\phi(L) C_{1,i}^{y,-j} A_i M(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r}\phi(\lambda_k) C_{1,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\ + \left[\frac{\sum_{i=1}^{v}\phi(L) C_{1,i}^{f,-j} f_i(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r}\phi(\lambda_k) C_{1,i}^{y,-j} f_i(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\ + \left[\frac{C_{1,i}^{g,-j} g(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{1,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \Bigg) V^{-1} B(L)^{-1} x_t \\ \vdots \\ \sum_{j=1}^{q}\sum_{\ell=0}^{j-1} \ell!\, L^{\ell-j} \Bigg(\left[\frac{\sum_{i=1}^{r}\phi(L) C_{r,i}^{y,-j} A_i M(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r}\phi(\lambda_k) C_{r,i}^{y,-j} A_i M(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\ + \left[\frac{\sum_{i=1}^{v}\phi(L) C_{r,i}^{f,-j} f_i(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{\sum_{i=1}^{r}\phi(\lambda_k) C_{r,i}^{y,-j} f_i(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \\ + \left[\frac{C_{r,i}^{g,-j} g(L) G(L)}{\Pi_{k=1}^{u}(L-\lambda_k)} - \sum_{k=1}^{u} \frac{C_{r,i}^{g,-j} g(\lambda_k) G(\lambda_k)}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}\right]_0^{(\ell)} \Bigg) V^{-1} B(L)^{-1} x_t \end{bmatrix}$$

This has to be true for all the possible realizations of $\{x_t\}$. By Riesz-Fischer Theorem, it is equivalent to the following system of functional equations

$$T(z)\phi(z) = D\left[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\right]$$

where $T(z)$ and $D\left[z, \{\phi(\lambda_k)\}_{k=1}^{u}, \{\phi^{(j)}(0)\}_{j=0}^{q}\right]$ are defined in equation (3.21) and (3.22), respectively.

By the Riesz-Fisher Theorem, there exists $\phi(L)$ that solves model (3.8) if and only if there exists a vector analytic function $\phi(z)$ that solves equations (3.20). □

A.8 Proof of Lemma 3.3

Proof. First note that $D\left[z, \{\phi(\lambda_k)\}_{k=1}^u, \{\phi^{(j)}(0)\}_{j=0}^q\right]$ is linear in constants $\{\{\phi(\lambda_k)\}_{k=1}^u, \{\phi^{(j)}(0)\}_{j=0}^q\}$. As a result, we can arrange $D\left[z, \{\phi(\lambda_k)\}_{k=1}^u, \{\phi^{(j)}(0)\}_{j=0}^q\right]$ to obtain

$$D\left[z, \{\phi(\lambda_k)\}_{k=1}^u, \{\phi^{(j)}(0)\}_{j=0}^q\right] \quad = \quad \widehat{D}_1(z)\begin{bmatrix}\phi(\lambda_1) & \dots & \phi(\lambda_u) & \phi^0(0) & \dots & \phi^q(0)\end{bmatrix}' + D_2(z)$$

Let $N_c = wu + w(q+1)$ denote the length of the vector $\begin{bmatrix}\phi(\lambda_1) & \dots & \phi(\lambda_u) & \phi^0(0) & \dots & \phi^q(0)\end{bmatrix}'$. Therefore, $\widehat{D}_1(z)$ is a $w \times N_c$ matrix. Let N_1 denote the column rank of $\widehat{D}_1(z)$. It follows that there exists N_1 vectors from $\widehat{D}_1(z)$ that consists a basis of $\widehat{D}_1(z)$. Denote these N_1 vectors as $D_1(z)$. Therefore, there exists a constant matrix Λ of dimension $N_1 \times N_c$, such that

$$\widehat{D}_1(z) = D_1(z)\Lambda\begin{bmatrix}\phi(\lambda_1) & \dots & \phi(\lambda_u) & \phi^0(0) & \dots & \phi^q(0)\end{bmatrix}'$$

Let $\psi \equiv \Lambda\begin{bmatrix}\phi(\lambda_1) & \dots & \phi(\lambda_u) & \phi^0(0) & \dots & \phi^q(0)\end{bmatrix}'$ completes the proof. □

A.9 Proof of Theorem 3

Proof. By Cramer's rule, the i-th element of $\phi(z)$ that solves equation (3.20) is given by

$$\phi_i(z) = \frac{\det\begin{bmatrix}T_1(z) & \dots & T_{i-1}(z) & D_1(z)\psi + D_2(z) & T_{i+1}(z)\dots & \dots & T_w(z)\end{bmatrix}}{\det\begin{bmatrix}T(z)\end{bmatrix}}$$

By Proposition 3.2, Proposition 3.1, and the assumption on model (3.8), the functions in $T(z)$, $D_1(z)$, and $D_2(z)$ 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.

In principle, the poles of $\phi_i(z)$ are either the roots of $\det[T(z)]$, i.e., $\{\vartheta_i, \dots, \vartheta_{N_2}\}$, or the poles of

$$\widehat{\phi}_i(z) \equiv \begin{bmatrix}T_1(z) & \dots & T_{i-1}(z) & D_1(z)\psi + D_2(z) & T_{i+1}(z)\dots & \dots & T_w(z)\end{bmatrix}. \tag{A.9}$$

By construction, the only poles of $\widehat{\phi}_i(z)$ are $\{\lambda_k\}_{k=1}^u$ and 0. However, $\{\lambda_k\}_{k=1}^u$ and 0 cannot be poles of $\phi_i(z)$ because these poles are generated from forming expectations using the Wiener-Hopf prediction formula, and by Proposition 3.1, these poles are already eliminated by $\left\{\{\phi(\lambda_k)\}_{k=1}^u, \{\phi^{(j)}(0)\right\}$.

Consider the inside roots of $\det[T(z)]$. For any ϑ_i, it is always possible to find ℓ_i such that $T_{\ell_i}(\vartheta_i)$ is a linear combination of $\left\{T_1(\vartheta_i), \dots, T_{\ell_i-1}(\vartheta_i),\ T_{\ell_i+1}(\vartheta_i), \dots, T_w(\vartheta_i)\right\}$. That is

$$T_{\ell_i}(\vartheta_i) = \sum_{k\neq\ell_i}\varphi_k^i T_k(\vartheta_i) \tag{A.10}$$

Suppose

$$\det\begin{bmatrix}D_1(\vartheta_1)\psi + D_2(\vartheta_1) & T_1(\vartheta_i) & \dots & T_{\ell_i-1}(\vartheta_i) & T_{\ell_i+1}(\vartheta_i) & \dots & T_w(\vartheta_i)\end{bmatrix} = 0 \tag{A.11}$$

Then for any $j \in \{1, \dots, \ell_i - 1, \ell_i + 1, \dots, w\}$, we have

$$\begin{aligned}
& \det\begin{bmatrix}T_1(\vartheta_i) & \dots & \overbrace{D_1(\vartheta_i)\psi + D_2(\vartheta_i)}^{j-th\ column} & \dots & \overbrace{T_{\ell_i}(\vartheta_i)}^{\ell_i-th\ column} & \dots & T_w(\vartheta_i)\end{bmatrix} \\
= & \sum_{k\neq\ell_i} \det\begin{bmatrix}T_1(\vartheta_i) & \dots & \overbrace{D_1(\vartheta_i)\psi + D_2(\vartheta_i)}^{j-th\ column} & \dots & \overbrace{\varphi_k^i T_k(\vartheta_i)}^{\ell_i-th\ column} & \dots & T_w(\vartheta_i)\end{bmatrix} \\
= & \ 0.
\end{aligned}$$

This implies that if equation (A.11) holds, for $j \in \{1, \dots, w\}$, ϑ_i is the root of the determinant

$$\det\begin{bmatrix}T_1(\vartheta_i) & \dots & \overset{j}{D_1(\vartheta_1)\psi + D_2(\vartheta_1)} & \dots & T_w(\vartheta_i)\end{bmatrix}$$

Consequently, ϑ_i cannot be a pole of $\phi(z)$. Now consider the following problem,

$$U_1\psi + U_2 \equiv \begin{bmatrix}\det\begin{bmatrix}D_1(\vartheta_1)\psi + D_2(\vartheta_1) & T_1(\vartheta_1) & \dots & T_{\ell_1-1}(\vartheta_1) & T_{\ell_1+1}(\vartheta_1) & \dots & T_w(\vartheta_1)\end{bmatrix} \\ \vdots \qquad\qquad \vdots \qquad \dots \qquad \vdots \qquad\qquad \vdots \qquad \dots \qquad \vdots \\ \det\begin{bmatrix}D_1(\vartheta_{N_2})\psi + D_2(\vartheta_{N_2}) & T_1(\vartheta_{N_{N_2}}) & \dots & T_{\ell_{N_2}-1}(\vartheta_{N_2}) & T_{\ell_{N_2}+1}(\vartheta_{N_2}) & \dots & T_w(\vartheta_{N_2})\end{bmatrix}\end{bmatrix} \tag{A.12}$$

If there exists ψ such that

$$U_1\psi + U_2 = 0 \tag{A.13}$$

Then $\{\vartheta_i\}_{i=1}^{N_2}$ are not poles of $\phi(z)$.

1. If $N_1 < N_2$, then there are more equations than unknowns. There does not exist ψ such that equation (A.9) holds. As a result, there is no solution to (3.20).
2. If $N_1 = N_2 = \text{rank}(U_2)$, then there exists a unique ψ that solves (A.9). Therefore, $\{\vartheta_i\}_{i=1}^{N_2}$ are not poles of $\phi(z)$.
3. If $N_1 > N_2$ or $N_1 = N_2 > \text{rank}(U_2)$, there are infinite solutions to (A.9). As a result, there are infinite number of analytic functions $\phi(z)$ that solves (3.20).

□

A.10 Proof of Theorem 2

Proof. By Theorem 3, if there exists a solution to (3.8), for $i \in \{1, \dots, w\}$, $\phi_i(z)$ is a rational function with finite degree. Therefore, $y_t = h(L)x_t$ can be written as (3.9). By Lemma 3.1, there exists a state space representation of $y_t = h(L)x_t$, which is given by

$$z_{t+1} = Fz_t + Qx_t \tag{A.14}$$

$$y_t = HQx_t + HFz_t \tag{A.15}$$

where F, Q and H are given by (A.1), (A.3), and (A.4) respectively. Define

$$\Gamma_x = HQ \tag{A.16}$$

$$\Gamma_z = HF \tag{A.17}$$

$$\Upsilon_x = Q \tag{A.18}$$

$$\Upsilon_z = F, \tag{A.19}$$

and we obtain the finite-state representation. Note that the eigenvalues of Γ_z all lie inside the unit circle. The law of motion of z_t

$$z_{t+1} = \Upsilon_x x_t + \Upsilon_z z_t \tag{A.20}$$

can be written as

$$z_{t+1} = (I - \Upsilon_z L)^{-1} \Upsilon_x x_t \tag{A.21}$$

Therefore, given $\{x_t\}_{t=-\infty}^{-1}$,

$$z_0 = (I - \Upsilon_z L)^{-1} \Upsilon_x x_{-1} \tag{A.22}$$

□

A.11 More on solution in innovation form

We follow the same procedure as the signal form to define the solution in innovation form. Here, we use similar notations as the signal form to make them comparable to each other, but it should be clear that they may stand for different objects.

Choice variable The policy rule we want to solve is $y_t = [y_{it}, \dots, y_{rt}]'$, where

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{rt} \end{bmatrix} = \begin{bmatrix} d_{11}(L) & \dots & h_{1m}(L) \\ \vdots & \dots & \vdots \\ h_{r1}(L) & \dots & h_{rm}(L) \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} = d(L) s_t. \tag{A.23}$$

We assume that each element in $d(L)$ has an infinite MA representation. More compactly, define

$$\phi(L) \equiv \begin{bmatrix} d_{11}(L) & \dots & d_{1m}(L) & \dots & d_{r1}(L) & \dots & d_{rm}(L) \end{bmatrix}. \tag{A.24}$$

Endogenous variables related to other agents' actions Let $f_t = [f_{it}, \dots, f_{vt}]'$ denote the endogenous variables chosen by other agents. They are related to the policy rule $\phi(L)$ and the driving shocks s_t in the following way

$$f_{it} = \phi(L) f^i(L) s_t = \phi(L) \begin{bmatrix} f^i_{11}(L) & \dots & f^i_{1m}(L) \\ \vdots & \dots & \vdots \\ f^i_{w1}(L) & \dots & f^i_{wm}(L) \end{bmatrix} \begin{bmatrix} s_{1t} \\ \vdots \\ s_{mt} \end{bmatrix} \tag{A.25}$$

Here, each $f_i(L)$ is a $w \times m$ matrix in the lag operator L. We assume that all the elements in $f_i(L)$ are finite rational functions in L and do not contain negative powers of L in expansion.

Exogenous variables This part is the same as the signal form exposition.

General model This part is the same as signal form.

Definition A.1. *A solution to model (3.8) (or an equilibrium) in innovation form is a vector of lag polynomials $\phi(L)$ such that*

1. *For each $i \in \{1, \dots, w\}$, $\phi_i(L)$ has an infinite MA representation*

$$\phi_i(L) = \sum_{k=0}^{\infty} \phi_{ik} L^k,$$

with $\sum_{k=0}^{\infty} \phi_{ik} < \infty$.

2. *For all possible realizations of $\{s_t\}$,*

$$y_t = \phi(L) \begin{bmatrix} A_1 & \dots & A_r \end{bmatrix}' x_t = d(L) s_t$$

satisfies equation (3.8).

A.12 Proof of Theorem 4

Proof. Suppose there exists a solution in signal form

$$y_t = h(L) x_t$$

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

$$y_t = h(L) M(L) s_t.$$

Because $y_t = h(L)x_t$ satisfies model (3.8), $y_t = h(L)M(L)s_t$ also satisfies model (3.8). Reversely, suppose there exists a solution in innovation form

$$y_t = d(L) s_t.$$

We can rearrange model (3.8) such that

$$y_t = -\left(\sum_{j=0}^{p} C^{y,j} L^j\right)^{-1} \left(\sum_{j=0}^{p} \mathbb{E}\left[C^{f,j} L^j f_t + C^{g,j} L^j g_t \middle| x^t\right] + \sum_{j=1}^{q} \mathbb{E}\left[C^{y,-j} L^{-j} y_t + C^{f,-j} L^{-j} f_t + C^{g,-j} L^{-j} g_t \middle| x^t\right]\right) \quad \text{(A.26)}$$

Note that $\left(\sum_{j=0}^{p} C^{y,j} L^j\right)$ has to be invertible. Otherwise, y_t is not co-variance stationary, which contradicts to the assumption that $y_t = d(L)s_t$ is a solution to the model. Therefore, $\{y_t\} \subset \mathcal{H}_t^x$ and $\{d(L)s_t\} \subset \mathcal{H}_t^x$. By Proposition 3.1, it follows that

$$y_t = d(L)s_t = \mathbb{E}[d(L)s_t | x^t] = \left(d(L)M'(L^{-1})\rho_{xx}(L)^{-1} - \sum_{k=1}^{u} \frac{d(\lambda_k)\lambda^k G(\lambda_k) V^{-1} B(L)^{-1}}{(L - \lambda_k)\Pi_{\tau \neq k}(\lambda_k - \lambda_\tau)}\right) x_t$$

Defining

$$h(L) = d(L)M'(L^{-1})\rho_{xx}(L)^{-1} - \sum_{k=1}^{u} \frac{d(\lambda_k)\lambda^k G(\lambda_k)V^{-1}B(L)^{-1}}{(L-\lambda_k)\Pi_{\tau\neq k}(\lambda_k - \lambda_\tau)}$$

gives us the signal form solution. □

A.13 Proof of Proposition 4.1

Proof. Consider the state-space representation of the signal process. The state equation is

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

The observation equation is

$$x_{it} = \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \xi_t + \begin{bmatrix} \epsilon_{it} \\ u_{it} \end{bmatrix}.$$

By the Canonical Factorization Theorem, the Wold representation is

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

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

where $\tau_1 = \sigma_\epsilon^2$ and $\tau_2 = \sigma_u^2$, and

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

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

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

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

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

The model requires that

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

which leads to

$$\begin{aligned} & h_1(L)x_{it}^1 + h_2(L)x_{it}^2 \\ = & \begin{bmatrix} \frac{1}{1-\lambda L}\frac{\lambda}{(1-\rho\lambda)\rho\tau_1} \\ \frac{1}{1-\lambda L}\frac{\lambda}{(1-\rho\lambda)\rho\tau_2} \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} \\ + & \alpha \begin{bmatrix} \frac{\lambda}{\rho\tau_1}\frac{L}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \\ \frac{\lambda}{\rho\tau_2}\frac{L}{(1-\lambda L)(L-\lambda)}h_1(L) - \frac{\lambda^2}{\rho\tau_2}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_1(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} \\ + & \alpha \begin{bmatrix} \frac{\lambda}{\rho\tau_1}\frac{L}{(1-\lambda L)(L-\lambda)}h_2(L) - \frac{\lambda^2}{\rho\tau_1}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \\ \frac{\lambda}{\rho\tau_2}\frac{L}{(1-\lambda L)(L-\lambda)}h_2(L) - \frac{\lambda^2}{\rho\tau_2}\frac{1}{1-\rho\lambda}\frac{1-\rho L}{(1-\lambda L)(L-\lambda)}h_2(\lambda) \end{bmatrix}' \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} \end{aligned}$$

By the Riesz-Fisher Theorem, we can transform it into the following problem

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

where $h(\lambda) = h_1(\lambda) + h_2(\lambda)$, and

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

Note that

$$\det C(z) = \frac{-\lambda\left[z^2 - \left(\frac{1}{\lambda} + \lambda - \frac{\alpha(\tau_1+\tau_2)}{\rho\tau_1\tau_2}\right)z + 1\right]}{(1-\lambda z)(z-\lambda)} = \frac{\frac{\lambda}{\vartheta}(z-\vartheta)(1-\vartheta z)}{(1-\lambda z)(z-\lambda)}$$

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,

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

The numerator is

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

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,

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

and similarly,

$$h_2(z) = \frac{\vartheta}{\rho\tau_2(1-\rho\vartheta)} \frac{1}{1-\vartheta z}$$

□

A.14 Proof of Proposition 4.3

Proof. The signal process and the Wold representation is the same as the proof A.13. The difference is when assuming $y_{it} = h_1(L)x_{it}^1 + h_2(L)x_{it}^2$, the aggregate y_t becomes

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

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

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

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 $h(\lambda) = h_2(\lambda)$,[14] and

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

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

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

[14]It can be verified that $h_1(\lambda)$ does not show up by using the property of λ.

Using Cramer's rule,

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

The numerator is

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

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,

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

□

A.15 Proof of Proposition 5.1

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

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

Given ϕ, let the signal process be

$$x_{it}^1 = \xi_t + \epsilon_{it},$$
$$x_{it}^2 = \phi_3(L)\eta_t + u_{it}.$$

Then the signal process is well defined. let

$$y_{it} = \phi_1(L)\epsilon_{it} + \phi_2(L)u_{it} + \phi_3(L)\eta_t,$$

and the optimal linear forecast is given by

$$\mathbb{E}_{it}[y_t] = \widehat{\phi}_1(L)\epsilon_{it} + \widehat{\phi}_2(L)u_{it} + \widehat{\phi}_3(L)\eta_t$$

If $y_{it} = \mathbb{E}_{it}[\xi_t] + \alpha \mathbb{E}_{it}[y_{jt}]$, then ϕ and Φ is an equilibrium.

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

$$\mathcal{T}(\phi) = \mathcal{T}(\{\phi_1, \phi_2, \phi_3\}) = \{\alpha\widehat{\phi}_1, \alpha\widehat{\phi}_2, \alpha\widehat{\phi}_3\}$$

The equilibrium is a fixed point of the operator $\mathcal{T}$. The proof of the contraction mapping is the same as the proof of Proposition 2.1. The modification is that the expectation will be conditional on the signal process that depends on ϕ. □

A.16 Proof of Theorem 5

Proof. Suppose the equilibrium policy rule $\phi = \{\phi_a, \phi_1, \phi_2, \phi_3\}$ allows a finite ARMA representation. Assume that

$$\phi_3(L) = \sigma_y \frac{\Pi_{k=1}^{q}(1 + \theta_k L)}{\Pi_{k=1}^{p}(1 - \rho_k L)}$$

where σ_y is a constant .

The signal process is then given by

$$\begin{aligned} x_{it}^1 &= \xi_t + \epsilon_{it} \\ x_{it}^2 &= y_t + u_{it} \end{aligned}$$

where

$$\begin{aligned} \xi_t &= \frac{\Pi_{k=1}^{n}(1 + \kappa_k L)}{\Pi_{k=1}^{m}(1 - \zeta_k L)}\eta_t \\ y_t &= \sigma_y \frac{\Pi_{k=1}^{q}(1 + \theta_k L)}{\Pi_{k=1}^{p}(1 - \rho_k L)}\eta_t \end{aligned}$$

The signal process can be rewritten as

$$x_{it} = \begin{bmatrix} x_{it}^1 \\ x_{it}^2 \end{bmatrix} = \begin{bmatrix} \sigma_\epsilon & 0 & \frac{\Pi_{k=1}^{n}(1+\kappa_k L)}{\Pi_{k=1}^{m}(1-\zeta_k L)} \\ 0 & \sigma_u & \sigma_y \frac{\Pi_{k=1}^{q}(1+\theta_k L)}{\Pi_{k=1}^{p}(1-\rho_k L)} \end{bmatrix} \begin{bmatrix} \widehat{a}_{m(i,t)} \\ \widehat{u}_{it} \\ \widehat{\eta}_t \end{bmatrix} = \widehat{M}(L)\widehat{s}_{it}$$

By the 3.4 Theorem, we can find the canonical factorization

$$\rho_{XX}(z) = \widehat{M}(z)\widehat{M}'(z^{-1}) = B(z)VB'(z^{-1})$$

Let $u_1 = \max\{n, m\}$, $u_2 = \max\{p, q\}$, $u = u_1 + u_2$, we have

$$B(z)^{-1} = \begin{bmatrix} \frac{b_1 \Pi_{k=1}^{m}(1-\zeta_k z)\Pi_{k=1}^{u_2}(1-r_k z)}{\Pi_{k=1}^{u}(1-t_k z)} & \frac{b_2 z \Pi_{k=1}^{p}(1-\rho_k z)\Pi_{k=u_2+1}^{u-1}(1-r_k z)}{\Pi_{k=1}^{u}(1-t_k z)} \\ \frac{b_3 z \Pi_{k=1}^{m}(1-\zeta_k z)\Pi_{k=u}^{u+u_2-2}(1-r_k z)}{\Pi_{k=1}^{u}(1-t_k z)} & \frac{b_4 \Pi_{k=1}^{p}(1-\rho_k z)\Pi_{k=u+u_2-1}^{2u-2}(1-r_k z)}{\Pi_{k=1}^{u}(1-t_k z)} \end{bmatrix}$$

$$B'(z^{-1})^{-1} = \begin{bmatrix} \frac{b_1 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=1}^{u_2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} & \frac{b_3 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=u}^{u+u_2-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} \\ \frac{b_2 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u_2+1}^{u-1}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} & \frac{b_4 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u+u_2-1}^{2u-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} \end{bmatrix}$$

and

$$V^{-1} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix}$$

Here, $\{t_k\}_{k=1}^{n}$ are eigenvalues of $F - FKH$ which are within the unit circle. $\{r_k\}_{k=1}^{2n-2}$, $\{b_k\}_{k=1}^{4}$ and $\{v_k\}_{k=1}^{4}$ are functions of the underlying parameters of $\widehat{M}(z)$.

Applying the Wiener-Hopf prediction formula, suppose we want to predict a random variable f_t, with

$$f_t = \begin{bmatrix} \psi_1(L) & \psi_2(L) & \psi_3(L) \end{bmatrix} \widehat{s}_{it}$$

where $\psi_1(L), \psi_2(L), \psi_3(L)$ do not contain negative powers. The prediction is

$$\mathbb{E}[f_t | x_i^t] = \left[\rho_{fx}(L) B'(L^{-1})^{-1}\right]_+ V^{-1} B(L)^{-1} M(L) s_{it} \tag{A.27}$$

It follows that

$$\begin{aligned} &\rho_{fx}(z) B'(z^{-1})^{-1} \\ = &\begin{bmatrix} \sigma_a \psi_1(z) + \frac{\psi_3(z) z^{m-n} \Pi_{k=1}^{n}(z+\kappa_k)}{\Pi_{k=1}^{m}(z-\zeta_k)} \\ \sigma_u \psi_2(z) + \frac{\psi_3(z) \sigma_\pi z^{p-q} \Pi_{k=1}^{q}(z+\theta_k)}{\Pi_{k=1}^{p}(z-\rho_k)} \end{bmatrix}' \begin{bmatrix} \frac{b_1 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=1}^{u_2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} & \frac{b_3 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=u}^{u+u_2-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} \\ \frac{b_2 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u_2+1}^{u-1}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} & \frac{b_4 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u+u_2-1}^{2u-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} \end{bmatrix} \\ = &\begin{bmatrix} \frac{\sigma_a \psi_1(z) b_1 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=1}^{u_2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} + \frac{\sigma_u \psi_2(z) b_2 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u_2+1}^{u-1}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} + \frac{\psi_3(z) e_1(z)}{\Pi_{k=1}^{u}(z-t_k)} \\ \frac{\sigma_a \psi_1(z) b_3 z^{u_1-m} \Pi_{k=1}^{m}(z-\zeta_k)\Pi_{k=u}^{u+u_2-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} + \frac{\sigma_u \psi_2(z) b_4 z^{u_2-p} \Pi_{k=1}^{p}(z-\rho_k)\Pi_{k=u+u_2-1}^{2u-2}(z-r_k)}{\Pi_{k=1}^{u}(z-t_k)} + \frac{\psi_3(z) e_2(z)}{\Pi_{k=1}^{u}(z-t_k)} \end{bmatrix}' \end{aligned}$$

where

$$e_1(z) = b_1 z^{u_1-n} \Pi_{k=1}^{n}(z+\kappa_k)\Pi_{k=1}^{u_2}(z-r_k) + b_2 \sigma_y z^{u_2-q} \Pi_{k=1}^{q}(z+\theta_k)\Pi_{k=u_2+1}^{u-1}(z-r_k)$$

$$e_2(z) = b_3 z^{u_1-n} \Pi_{k=1}^{n}(z+\kappa_k)\Pi_{k=u}^{u+u_2-2}(z-r_k) + b_4 \sigma \pi z^{u_2-q} \Pi_{k=1}^{q}(z+\theta_k)\Pi_{k=u+u_2-1}^{2u-2}(z-r_k)$$

Also,

$$V^{-1}B(z)^{-1} = \frac{1}{\Pi_{k=1}^{u}(1-t_k z)} \begin{bmatrix} \Pi_{k=1}^{m}(1-\zeta_k z)D_{11}(z) & \Pi_{k=1}^{p}(1-\rho_k z)D_{12}(z) \\ \Pi_{k=1}^{m}(1-\zeta_k z)D_{21}(z) & \Pi_{k=1}^{p}(1-\rho_k z)D_{22}(z) \end{bmatrix}$$

where

$$D(z) = \begin{bmatrix} v_1 b_1 \Pi_{k=1}^{u_2}(1-r_k z) + v_2 b_3 z \Pi_{k=u}^{u+u_2-2}(1-r_k z) & v_1 b_2 \Pi_{k=u_2+1}^{u-1}(1-r_k z) + v_2 b_4 \Pi_{k=u+u_2-1}^{2u-2}(1-r_k z) \\ v_3 b_1 \Pi_{k=1}^{u_2}(1-r_k z) + v_4 b_3 z \Pi_{k=u}^{u+u_2-2}(1-r_k z) & v_3 b_2 \Pi_{k=u_2+1}^{u-1}(1-r_k z) + v_4 b_4 \Pi_{k=u+u_2-1}^{2u-2}(1-r_k z) \end{bmatrix}$$

$$\left[\frac{g(z)}{(z-\lambda_1)\cdots(z-\lambda_u)}\right]_+ = \frac{g(z)}{(z-\lambda_1)\cdots(z-\lambda_u)} - \sum_{k=1}^{u} \frac{g(\lambda_k)}{(z-\lambda_k)\Pi_{\tau\neq k}(\lambda_k-\lambda_\tau)}$$

To predict of y_t, using equation A.27, the components are

$$\begin{aligned} \mathbb{E}_{it}[\xi_t] &= \frac{1}{\Pi_{k=1}^{u}(1-t_k L)} \begin{bmatrix} \frac{\Pi_{\ell=1}^{n}(1+\kappa_\ell L)e_1(L)}{\Pi_{\ell=1}^{m}(1-\zeta_\ell L)\Pi_{k=1}^{u}(L-t_k)} - \sum_{k=1}^{u} \frac{\Pi_{\ell=1}^{n}(1+\kappa_\ell t_k)e_1(t_k)}{\Pi_{\ell=1}^{m}(1-rho_\ell t_k)(L-t_k)\Pi_{\tau\neq k}(t_k-t_\tau)} \\ \\ \frac{\Pi_{\ell=1}^{n}(1+\kappa_\ell L)e_2(L)}{\Pi_{\ell=1}^{m}(1-\zeta_\ell L)\Pi_{k=1}^{u}(L-t_k)} - \sum_{k=1}^{u} \frac{\Pi_{\ell=1}^{n}(1+\kappa_\ell t_k)e_2(t_k)}{\Pi_{\ell=1}^{m}(1-rho_\ell t_k)(L-t_k)\Pi_{\tau\neq k}(t_k-t_\tau)} \end{bmatrix}' V^{-1}B(L)^{-1}M(L)s_{it}, \\ &= \frac{1}{\Pi_{k=1}^{u}(1-t_k L)} \begin{bmatrix} g_1(L) & g_2(L) & g_3(L) \end{bmatrix} s_{it} \end{aligned}$$

where $g_1(L)$, $g_2(L)$, and $g_3(L)$ do not contain negative powers and are independent of the equilibrium policy rule ϕ.

$$\mathbb{E}_{it}[\phi_3(L)\eta_t] =$$

$$\begin{bmatrix} \Pi_{k=1}^{m}(1-\zeta_k L)\left\{\frac{e_1(L)D_1(L)+e_2(L)D_3(L)}{\Pi_{k=1}^{u}(1-t_k L)(L-t_k)}\phi_3(L) - \sum_{\tau=1}^{u} \frac{e_1(t_\tau)D_1(L)+e_2(t_\tau)D_3(L)}{(L-t_\tau)\Pi_{k=1}^{u}(1-t_k L)\Pi_{k\neq\tau}(t_\tau-t_k)}\phi_3(t_\tau)\right\} \\ \\ \Pi_{k=1}^{p}(1-\rho_k L)\left\{\frac{e_1(L)D_2(L)+e_2(L)D_4(L)}{\Pi_{k=1}^{u}(1-t_k L)(L-t_k)}\phi_3(L) - \sum_{\tau=1}^{u} \frac{e_1(t_\tau)D_2(L)+e_2(t_\tau)D_4(L)}{(L-t_\tau)\Pi_{k=1}^{u}(1-t_k L)\Pi_{k\neq\tau}(t_\tau-t_k)}\phi_3(t_\tau)\right\} \\ \\ \frac{e_1(L)[\Pi_{k=1}^{n}(1+\kappa_k L)D_1(L)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k L)D_2(L)]+e_2(L)[\Pi_{k=1}^{n}(1+\kappa_k L)D_3(L)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k L)D_4(L)]}{\Pi_{k=1}^{u}(1-t_k L)(L-t_k)}\phi_3(L) \\ - \sum_{\tau=1}^{n} \frac{e_1(t_\tau)[\Pi_{k=1}^{n}(1+\kappa_k L)D_1(L)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k L)D_2(L)]+e_2(t_\tau)[\Pi_{k=1}^{n}(1+\kappa_k L)D_3(L)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k L)D_4(L)]}{(L-t_\tau)\Pi_{k=1}^{u}(1-t_k L)\Pi_{k\neq\tau}(t_\tau-t_k)}\phi_3(t_\tau) \end{bmatrix}' \begin{bmatrix} a_{m(i,t)} \\ u_{it} \\ \eta_t \end{bmatrix}.$$

The equilibrium condition is

$$\phi_1(L)\epsilon_{m(i,t)} + \phi_2(L)u_{it} + \phi_3(L)\eta_t = \mathbb{E}_{it}[\xi_t] + \alpha\, \mathbb{E}_{it}[\phi_3(L)\eta_t]$$

By the Riesz-Fisher Theorem, we can solve the corresponding analytical functions. Particularly, if we can solve for $\phi_3(z)$, then $\phi_1(z)$ and $\phi_2(z)$ can be solved easily. $\phi_3(z)$ needs to satisfy the following condition

$$\begin{aligned}\phi_3(z) &= \frac{g_3(z)}{\Pi_{k=1}^n(1-t_kz)} \\ &+ \alpha\frac{e_1(z)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_2(z)]+e_2(z)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma\pi\Pi_{k=1}^q(1+\theta_kz)D_4(z)]}{\Pi_{k=1}^u(1-t_kz)(z-t_k)}\phi_3(z) \\ &- \alpha\sum_{\tau=1}^{u}\frac{e_1(t_\tau)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_2(z)]+e_2(t_\tau)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_4(z)]}{(z-t_\tau)\Pi_{k=1}^u(1-t_kz)\Pi_{k\neq\tau}(t_\tau-t_k)}\phi_3(t_\tau)\end{aligned}$$

Multiplying $\Pi_{k=1}^u(1-t_kz)(z-t_k)$ to both sides leads to

$$\begin{aligned}&\Big\{\Pi_{k=1}^u(1-t_kz)(z-t_k) \\ &\quad-\alpha\Big(e_1(z)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kL)D_2(z)]+e_2(z)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_4(z)]\Big)\Big\}\phi_3(z) \\ &= \Pi_{k=1}^u(z-t_k)g_3(z) \\ &-\alpha\sum_{\tau=1}^{u}\frac{\Pi_{k\neq\tau}(z-t_k)\{e_1(t_\tau)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_2(z)]+e_2(t_\tau)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_4(z)]\}}{\Pi_{k\neq\tau}(t_\tau-t_k)}\phi_3(t_\tau)\end{aligned}$$

The denominator of $\phi_3(z)$ (before eliminating inside roots) is

$$\Pi_{k=1}^u(1-t_kz)(z-t_k)-\alpha\Big(e_1(z)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kL)D_2(z)]+e_2(z)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_4(z)]\Big) \tag{A.28}$$

If we can show that the roots of the denominator of $\phi_3(z)$ are always different from $\{\rho_k\}_{k=1}^p$, then the proof is done. Here we show that $\{\rho_k\}_{k=1}^p$ can be the roots of the denominator of $\phi_3(z)$ only if $\alpha = 1$.

Consider the non-casual prediction formula for random variable f_t, i.e., the prediction based on x_i^∞.

$$\begin{aligned}\mathbb{E}[f_t|x_i^\infty] &= \rho_{fx}(L)\rho_{xx}(L)^{-1}M(L)s_{it} && \text{(A.29)}\\ &= \rho_{fx}(L)[\widehat{M}(L)\widehat{M}'(L^{-1})]^{-1}M(L)s_{it} && \text{(A.30)}\\ &= \rho_{fx}(L)[B(L)VB'(L^{-1})]^{-1}M(L)s_{it} && \text{(A.31)}\end{aligned}$$

We focus on the third component of the prediction of $\eta_t = [0,0,1]s_{it}$. Using formula (A.31)

$$\begin{aligned}&\rho_{\eta x}(z)[B(z)VB'(z^{-1})]^{-1}M(z)\begin{bmatrix}0\\0\\1\end{bmatrix} \\ =\;& \frac{e_1(z)[\Pi_{k=1}^n(1+\kappa_kz)D_1(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_2(z)]+e_2(z)[\Pi_{k=1}^n(1+\kappa_kz)D_3(z)+\sigma_y\Pi_{k=1}^q(1+\theta_kz)D_4(z)]}{\Pi_{k=1}^u(1-t_kz)(z-t_k)}\end{aligned}$$

The equation above is the third row of the prediction formula (innovation form) for η_t when the prediction is based on x_i^∞, i.e., the non-casual prediction. This equation has an intimate link with equation (A.28).

Using formula (A.30), define $\tau_1 = \sigma_a^2, \tau_2 = \sigma_u^2$ and $\tau_3 = \sigma_y^2$. Also define $\Delta(z)$ as

$$\begin{aligned}\Delta(z) &= \tau_2 z^{m-n}\Pi_{k=1}^{n}(z+\kappa_k)(1+\kappa_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z) + \tau_1\tau_3 z^{p-q}\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{q}(z+\theta_k)(1+\theta_k z) \\ &\quad +\tau_1\tau_2\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z)\end{aligned}$$

We have

$$\begin{aligned}&\rho_{\eta x}(z)[\widehat{M}(z)\widehat{M}'(z^{-1})]^{-1}M(z)\begin{bmatrix}0\\0\\1\end{bmatrix} \\ = &\frac{\Delta(z) - \tau_1\tau_2\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z)}{\Delta(z)}.\end{aligned}$$

Because formula (A.31) equals formula (A.30), we have

$$\begin{aligned}&\Pi_{k=1}^{u}(1-t_k z)(z-t_k) - \left(e_1(z)[D_1(z)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k L)D_2(z)] + e_2(z)[D_3(z)+\sigma_y\Pi_{k=1}^{q}(1+\theta_k z)D_4(z)]\right) \\ \propto\ &\tau_1\tau_2\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z),\end{aligned}$$

where

$$\tau_1\tau_2\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z) = \Delta(z) - [\Delta(z) - \tau_1\tau_2\Pi_{k=1}^{m}(z-\zeta_k)(1-\zeta_k z)\Pi_{k=1}^{p}(z-\rho_k)(1-\rho_k z)].$$

Therefore, the denominator of $\phi_3(z)$ has roots $\{\rho_k\}_{k=1}^{p}$ only when $\alpha = 1$. If $\alpha \in (0,1)$, the roots of the denominator cannot include $\{\rho_k\}_{k=1}^{p}$.

□

A.17 Proof of Proposition 6.1

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

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

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

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

where

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

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

$$\mathbb{E}_{it}[y_{m(i,t)t}] = \mathbb{E}_{it}[h_a a_{m(i,t)} + h_1(L)(a_{m(m(i,t),t)} + \epsilon_{m(i,t)t}) + h_2(L)(u_{(m(i,t)t)} + \epsilon_{m(m(i,t),t)} + \xi_t)],$$

where

$$\begin{aligned}
\mathbb{E}_{it}[a_{m(i,t)}] &= \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2} x_{it}^1 \\
\mathbb{E}_{it}[a_{m(m(i,\tau),\tau)}] &= a_i \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[\epsilon_{m(i,\tau)\tau}] &= \frac{\sigma_\epsilon^2 v(1-\rho L)}{1-\lambda L}\widehat{x}_{it}^2 \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[u_{m(i,t)t}] &= 0 \\
\mathbb{E}_{it}[\epsilon_{m(m(i,\tau),\tau)}] &= \frac{\sigma_\epsilon^2}{\sigma_a^2 + \sigma_\epsilon^2} x_{it}^1 \quad \text{if } \tau = t, \text{ otherwise } 0 \\
\mathbb{E}_{it}[h_2(L)\xi_t] &= \left(\frac{vLh_2(L)}{(L-\lambda)(1-\lambda L)} - \frac{v\lambda(1-\rho L)h_2(\lambda))}{(1-\rho\lambda)(L-\lambda)(1-\lambda L)}\right)\widehat{x}_{it}^2.
\end{aligned}$$

The system is

$$\begin{aligned}
&h_a a_i + h_1(L)x_{it}^1 + h_2(L)\widehat{x}_{it}^2 \\
=&a_i + \alpha\Bigg[h_a \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2} x_{it}^1 + h_1(0)a_i + h_1(0)\frac{\sigma_\epsilon^2 v(1-\rho L)}{1-\lambda L}\widehat{x}_{it}^2 \\
&+ \left(\frac{vLh_2(L)}{(L-\lambda)(1-\lambda L)} - \frac{v\lambda(1-\rho L)h_2(\lambda))}{(1-\rho\lambda)(L-\lambda)(1-\lambda L)}\right)\widehat{x}_{it}^2 + h_2(0)\frac{\sigma_\epsilon^2}{\sigma_a^2 + \sigma_\epsilon^2} x_{it}^1\Bigg],
\end{aligned}$$

which leads to

$$\begin{aligned}
h_a &= 1 + \alpha h_1(0) \\
h_1(0) &= \alpha h_a \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2} + \alpha_1 h_2(0)\frac{\sigma_1^2}{\sigma_a^2 + \sigma_1^2} \\
h_2(z) &= \alpha h_1(0)\frac{\sigma_\epsilon^2 v(1-\rho z)}{1-\lambda z} + \alpha\left(\frac{vzh_2(z)}{(z-\lambda)(1-\lambda z)} - \frac{v\lambda(1-\rho z)h_2(\lambda)}{(1-\rho\lambda)(z-\lambda)(1-\lambda z)}\right).
\end{aligned}$$

The third equation can be written as

$$-\lambda(z-\vartheta)\left(z - \frac{1}{\vartheta}\right)h_2(z) = \alpha_1 h_1 \sigma_1^2 v(1-\rho z)(z-\lambda) - \alpha\frac{\sigma_\eta^2 v\lambda(1-\rho z)h_2(\lambda)}{(1-\rho\lambda)}$$

where

$$\vartheta = \frac{1}{2}\left[\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho(\sigma_\epsilon^2 + \sigma_u^2)} - \sqrt{\left(\frac{1}{\rho} + \rho + \frac{1-\alpha}{\rho(\sigma_\epsilon^2 + \sigma_u^2)}\right)^2 - 4}\right]. \tag{A.32}$$

Use $h_2(\lambda)$ to removes the inside root ϑ, we have

$$
\begin{aligned}
h_1(z) = h_1(0) &= \frac{\alpha}{1-\alpha^2 + \frac{\sigma_\epsilon^2}{\sigma_a^2}\left(1 - \alpha^2 \frac{\vartheta}{\rho}\frac{\sigma_\epsilon^2}{\sigma_\epsilon^2+\sigma_u^2}\right)} \\
h_a &= 1 + \alpha h_1(0) \\
h_2(z) &= \frac{\alpha\vartheta h_1(0)\sigma_\epsilon^2}{\rho(\sigma_\epsilon^2+\sigma_u^2)}\frac{1-\rho z}{1-\vartheta z}
\end{aligned}
$$

□