Analytic Policy Function Iteration

Zhao Han, Fei Tan and Jieran Wu

PII: S0022-0531(21)00212-X
DOI: https://doi.org/10.1016/j.jet.2021.105395
Reference: YJETH 105395

To appear in: *Journal of Economic Theory*

Received date: 29 April 2020
Revised date: 21 November 2021
Accepted date: 22 November 2021

Analytic Policy Function Iteration[†]

Zhao Han, Fei Tan, and Jieran Wu*

[First Version: December 20, 2019]
[This Version: November 24, 2021]

Abstract

We propose an approach to solving and analyzing linear rational expectations models with general information frictions. Our approach is built upon policy function iterations in the frequency domain. We develop the theoretical framework of this approach using rational approximation, analytic continuation, and discrete Fourier transform. Conditional expectations, which are difficult to evaluate in the time domain, can be calculated efficiently in the frequency domain. We provide the numerical implementation accompanied by a flexible object-oriented toolbox. We demonstrate the efficiency and accuracy of our method by studying four models in macroeconomics and finance that feature asymmetric information sets, endogenous signals, and higher-order expectations.

Keywords: Endogenous information; Policy function iteration; Frequency-domain methods; Higher-order expectations.

JEL Classification: C6, D8, E3, G1

[†]We thank the editor Guillermo Ordoñez, two anonymous referees, Ryan Chahrour, Zhen Huo, Kenneth Kasa, Siming Liu, Yulei Luo, Todd Walker, Eric Young, and Renbin Zhang as well as seminar participants at Zhejiang University and Shanghai University of Finance and Economics for helpful comments. We are also grateful for discussions with Lasha Ephremidze and Ilya Spitkovsky on spectral factorization methods. Z. Han acknowledges the financial support from the William & Mary Faculty Research Grant. F. Tan acknowledges the financial support from the Chaifetz School of Business summer research grant. J. Wu acknowledges the support from the Fundamental Research Funds for the Central Universities in China.

*Han: Department of Economics, William & Mary; Tan: Department of Economics, Chaifetz School of Business, Saint Louis University and Center for Economic Behavior and Decision-Making, Zhejiang University of Finance and Economics; Wu: Academy of Financial Research and School of Economics, Zhejiang University. Send correspondence to email: jw5ya@zju.edu.cn (J. Wu).

1 Introduction

Incorporating information frictions in dynamic general equilibrium models has produced fruitful results that shed light on a wide range of questions in macroeconomics and finance [Angeletos and Lian (2016)]. Despite the progress made so far, solving dynamic models with endogenous information frictions remains a daunting task. To address this challenge, this article proposes a frequency-domain method of solving and characterizing macroeconomic and finance models with endogenous information frictions.

The presence of endogenous information is inevitable in most macroeconomic models with rational expectations. Households and firms form expectations conditional on endogenous economic conditions, whose decisions, in turn, impact the state of the economy. In models of the financial market, learning from endogenous asset prices is also an indispensable ingredient as these prices aggregate information. Another notion of information endogeneity emerges in disciplined bounded rationality models, such as rational inattention models, where the choice of information structure is endogenous.

While endogenous information constitutes an essential ingredient of the general equilibrium (GE) feedback mechanism between expectations and economic outcomes that produces propagation, persistence, and volatility [e.g., Angeletos and Lian (2018), Chahrour and Gaballo (2019)], it also complicates the model solution significantly. In models with endogenous information, an information fixed point exists between agents' perceived law of motion of the economy (including endogenous variables such as prices) and the actual law of motion, based on agents' actions and conditional expectations using the endogenous signals. The underlying GE effect, operating through the lens of endogenous learning, creates equilibria that in general admit no finite-state representation.[1]

The complication is exacerbated in the environment of information heterogeneity. When an agent's view about economic fundamentals differs from that of other agents, the "forecasting the forecasts of others" problem gives rise to the role of higher-order expectations (HOEs) in shaping model dynamics. The recursion of HOEs implies that agents need to form an infinite order of expectations about what others believe when making decisions. When agents are not learning from endogenous variables, Huo and Takayama (2018) show that HOEs are tractable, and the model equilibrium permits a finite-state representation in the time domain. Therefore, the greatest challenge to solving and analyzing the model arises when information is both endogenous and heterogeneous.

The key contribution of this paper is to develop an analytic policy function iteration (APFI)

[1]Makarov and Rytchkov (2012) first illustrate this point in an asset pricing model. Models with endogenous information may admit finite-state representation when the signal structure is "square" with equal numbers of signals and innovations. The finite-state representation is defined in the time series sense that the equilibrium variables follow a finite-order VARMA process.

method to address the challenge directly. We use a simple asset pricing model of Singleton (1987) to demonstrate the basic idea of our method (Section 2). We circumvent the issue with infinite-state representation in the time domain by treating the frequency domain as the appropriate state space.[2] Our idea is straightforward and analogous to the classical method used to solve dynamic programming problems in the time domain. However, the mathematical foundation and numerical implementation of our approach differ substantially from standard policy function iterations.

We first provide three theorems that establish the theoretical foundation for the APFI method (Section 3). The first theorem characterizes the basis for functional approximations in the frequency domain. We use the set of rational functions to approximate the true equilibrium solution. The functional form of the solution within each iteration is known (as rational analytic functions). We show that any linear stationary equilibrium can be approximated arbitrarily well by a VARMA(p, q) process. The second theorem uses the theory of analytic continuation to construct the appropriate state-space grid in the real unit interval $(-1, 1)$. We apply the theory of the convergence of analytic functions to establish a convergence criterion for our algorithm. The third theorem constructs a fast and efficient method of computing conditional expectations in the frequency domain, which are typically hard to evaluate in the time domain. Specifically, we apply the discrete Fourier transform to compute conditional expectations under different information sets.

We then establish the baseline APFI algorithm along with numerical details that facilitate the computation (Section 4). Admittedly, frequency-domain methods have not been widely adopted by macroeconomists. To minimize the user's fixed cost, we provide a handy MATLAB-based, object-oriented toolbox called "z-Tran" that implements these procedures, serving as the paper's second contribution.[3] This toolbox encapsulates all required frequency-domain methods via a user-friendly interface. An applied user can quickly input the model's linearized equilibrium conditions into the canonical form of the baseline APFI algorithm and test the model implications. We also allow experienced users to call each routine in the toolbox independently and modify the baseline APFI algorithm whenever appropriate.

The existing frequency-domain method is powerful in deriving analytical characterization of the equilibrium [e.g., Whiteman (1983), Tan and Walker (2015), and Huo and Takayama (2018)]. However, it also faces limitations due to the cumbersome symbolic algebra involved in the procedure. The APFI framework circumvents these limitations. With numerical efficiency and stability, it extends the applicability of existing methods from characterizing small, illustrative

[2]The frequency-domain approach solves and analyzes model equilibrium in the space of analytic functions. Early contributions include Hansen and Sargent (1980), Whiteman (1983, 1985), and Taub (1989), and more recent developments include Kasa (2000), Kasa, Walker and Whiteman (2014), Acharya (2014), Tan and Walker (2015), Huo and Takayama (2018), Rondina and Walker (2018), Miao, Wu and Young (2021b), and Al-Sadoon (2018, 2020), among others.

[3]The toolbox is publicly available, free of charge, at https://github.com/econdojo/ztran.

models to solving large, quantitative dynamic stochastic general equilibrium (DSGE) models with general information frictions. The APFI framework is also flexible in the choice of information structure, as our canonical representation nests nearly all examples considered in the literature, including full information, imperfect (exogenous or endogenous) information, and heterogeneous (dispersed or hierarchical) information.

In related research, Huo and Takayama (2018) characterize the analytical solution to incomplete-information models by applying the state-space method to obtain the Wold fundamental representation in the frequency domain. While Huo and Takayama (2018) focus primarily on exogenous information, our APFI approach is designed mainly for models with endogenous information where the signal process per se does not admit an exact VARMA representation. It is also suitable for models whose explicit solutions from algebraic derivation become infeasible.[4] In this sense, our APFI framework complements the approach of Huo and Takayama (2018) and the two methods agree when information is exogenous.

Another influential work by Nimark (2017) develops a state-space method of solving dynamic models of dispersed information. His method truncates the (potentially) infinite-dimensional state vector of HOEs to a finite order and is widely adopted in the literature. One limitation of the truncation approach arises when agents in the model face asymmetric information frictions. That is, different groups of agents face ex-ante distinct information structures. In this case, the truncation strategy no longer works due to the explosion of cross-expectations among groups. Another limitation of the truncation approach emerges when the time series structure of the primitive model extends beyond the simple AR(1) recursion so that pinning down a suitable state space becomes tricky. In contrast, the APFI framework is not constrained by these limitations and works well for models with general information frictions.[5]

In summary, the APFI framework is particularly useful for solving dynamic models with endogenous signals and substantial information heterogeneity. We demonstrate the reliability and flexibility of our method by applying it to study three macroeconomic models (Section 5). The first example solves a prototypical New Keynesian DSGE model similar to Melosi (2017). We highlight the sensitivity of model solutions to incomplete-information firms' endogenous signals. We then augment this DSGE model with a fiscal sector that features primary surplus and gov-

[4]In the extension of the paper, Huo and Takayama (2018) also propose a numerical algorithm to solve the endogenous information case. Unlike our method that iterates on the policy function values, their method makes a conjecture of the parametric VARMA form and iterates on the VARMA coefficients, which requires lengthy algebra in evaluating the expectations and a non-trivial procedure for updating the VARMA orders.

[5]Our paper is also related to several recent papers that attempt to bridge the gap between incomplete-information models and their full-information counterparts by constructing an isomorphism between the two. The key ingredient in this approach is to introduce certain forms of distortions or wedges so that the incomplete-information economies are equivalent to the modified complete-information economies. Examples in this line of research include Chahrour and Ulbricht (2018) and Angeletos and Huo (2021). While these papers concentrate on the theoretical insight from the indirect mapping and its qualitative implications, our paper adopts a direct computational approach emphasizing the information endogeneity.

ernment debt and introduce incomplete-information households. We showcase four distinct fiscal effects of the primary surplus shock by allowing firms and households to observe differential, non-nested, and endogenous information. Our results provide new insight into how fiscal policy affects inflation.

The second example considers a HANK-type model of Angeletos and Huo (2021) with endogenous wealth distribution and incomplete information. We modify their original model by allowing different groups of households to be endowed with asymmetric, endogenous information sets. We examine how such asymmetric information frictions interact with the group heterogeneity and shape the model dynamics. To the best of our knowledge, the extension of introducing distinct, endogenous information frictions to different sectors (groups) of the economy is novel, which highlights the flexibility of our methodology.

Our baseline APFI algorithm requires an invertibility condition on the non-expectational block of the model system. Moreover, it cannot handle models with random walk dynamics. In the last application, we consider such a dispersed-information RBC model from Graham and Wright (2010). The model features non-stationarity and multiple equilibria, and it does not satisfy the invertibility condition. To circumvent the limitations of the baseline APFI algorithm, we design three extended APFI algorithms to solve this model. We also conduct a numerical experiment that compares the performance of our algorithms with the time-domain truncation method, which demonstrates the comparative advantages of the APFI framework in terms of accuracy, flexibility, and initial conjecture choice.

2 A Simple Model

We use a simple asset pricing model to illustrate the basic idea of our methodology. We also discuss its advantages vis-à-vis other popular frequency-domain and time-domain approaches.

2.1 Environment

The model environment follows Singleton (1987). There are two assets in the market: a stock with stochastic dividend payment d_t and a risk-free one-period bond with constant gross return $R > 1$. The stock is in zero fixed supply. A continuum of short-lived traders indexed by $i \in [0, 1]$ allocate their wealth optimally between these two assets to maximize their constant absolute risk aversion utilities. It is well-known that the equilibrium condition for the (cum-dividend) stock price p_t is given by

$$p_t = \beta \overline{\mathbb{E}}_t p_{t+1} + d_t \tag{2.1}$$

where $\overline{\mathbb{E}}_t(\cdot) = \int_0^1 \mathbb{E}_{i,t}(\cdot)di$ is the cross-sectional average expectation operator at time t and $\beta = 1/R$ defines the constant discount factor.

The market fundamental or the dividend is unobservable to traders when they trade. Suppose the exogenous dividend d_t is a covariance-stationary process driven by a persistent AR(1) shock and a transitory shock

$$d_t = e_t + \eta_t = \frac{1}{1-\rho\,\mathrm{L}}\varepsilon_t + \eta_t, \qquad \varepsilon_t \sim \mathbb{N}(0,\sigma_\varepsilon^2), \qquad \eta_t \sim \mathbb{N}(0,\sigma_\eta^2) \tag{2.2}$$

where L is the lag operator, $\mathrm{L}^k\,\varepsilon_t = \varepsilon_{t-k}$, and $0 < \rho < 1$. During each period every trader receives a private signal $s_{i,t}$ on the persistent fundamental innovation ε_t

$$s_{i,t} = \varepsilon_t + \nu_{i,t}, \qquad \nu_{i,t} \sim \mathbb{N}(0,\sigma_\nu^2) \tag{2.3}$$

The noise component $\nu_{i,t}$ is *i.i.d.* across traders i and over time t. These idiosyncratic noises are dispersed in the sense that

$$\int_0^1 \nu_{i,t}di \equiv 0 \tag{2.4}$$

The innovations $(\varepsilon_t, \nu_{i,t}, \eta_t)$ are uncorrelated at all leads and lags.

Trader i's information set is defined as $\Omega_{i,t} = s_i^t \vee p^t$, where $s_i^t \equiv \{s_{i,t}, s_{i,t-1}, \ldots\}$, $p^t \equiv \{p_t, p_{t-1}, \ldots\}$, and $\vee$ denotes the smallest closed subspace generated by the history of exogenous signals and endogenous prices. The conditional expectation operator $\mathbb{E}_{i,t}(\cdot)$ refers to trader i's individual expectation, i.e., $\mathbb{E}_{i,t}(\cdot) = \mathbb{E}[\cdot|\Omega_{i,t}]$. We restrict our attention to the symmetric equilibrium in which all traders behave according to the same linear decision rule.

2.2 Frequency-Domain Preliminaries

We characterize the linear covariance-stationary equilibrium price using polynomials in the lag operator L

$$p_t = \sum_{k=0}^{\infty} A_k\,\mathrm{L}^k\,\varepsilon_t + \sum_{k=0}^{\infty} B_k\,\mathrm{L}^k\,\eta_t \equiv A(\mathrm{L})\varepsilon_t + B(\mathrm{L})\eta_t, \tag{2.5}$$

where $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$ are square-summable sequences of coefficients, i.e., $\sum_{k=0}^{\infty}|A_k|^2 < \infty$ and $\sum_{k=0}^{\infty}|B_k|^2 < \infty$. The conjecture is valid as any causal covariance-stationary equilibrium process has a (one-sided) infinite-order moving average representation, i.e., an MA(∞) representation. The economic interpretation of $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$ is straightforward—they measure the k-period-ahead impulse response of the asset price to a one unit increase in the structural innovations, which must converge to zero over time by stationarity.

Solving for the infinite sequences of impulse responses $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$ in the time domain is a daunting task. Instead, we adopt the frequency-domain approach. The basic idea of frequency-domain methods is to transform the problem by replacing the lag operator L with a complex-valued variable z. We then solve an equivalent yet simpler problem of searching for the analytic functions $A(z)$ and $B(z)$ in the complex plane

$$A(z) = \sum_{k=0}^{\infty} A_k z^k, \qquad B(z) = \sum_{k=0}^{\infty} B_k z^k, \qquad z \in \mathbb{D} \tag{2.6}$$

where $\mathbb{D}$ denotes the open unit disk $\{z \in \mathbb{C} : |z| < 1\}$ and $\mathbb{C}$ denotes the complex plane. The functions $A(z)$ and $B(z)$ are called the z-transforms of $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$, and they encode the whole impulse response sequences. These functions completely summarize the covariogram and hence the second moment properties of $\{p_t\}$ via the covariance-generating function $S_p(z) = \sum_{k=-\infty}^{\infty} \mathbb{E}(p_t p_{t-k}) z^k = \sigma_\varepsilon^2 A(z) A\left(z^{-1}\right) + \sigma_\eta^2 B(z) B\left(z^{-1}\right)$.

Formally, the equivalence between the two representations is established by the Riesz-Fischer theorem, which states that there exists an isometrically isomorphic mapping from the space of one-sided, square-summable sequences to the Hardy space of analytic functions $\mathbf{H}^2(\mathbb{D})$.[6] Therefore, solving for the sequences $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$ in (2.5) amounts to solving for the functions $A(z)$ and $B(z)$ in $\mathbf{H}^2(\mathbb{D})$. We refer interested readers to Online Appendix S1 for more technical details.

In practice, the MA(∞) process (2.5) can be approximated arbitrarily well by an autoregressive moving-average process of finite orders p and q, i.e., ARMA(p, q). For example, if we shut down the transitory shock η_t, the equilibrium asset price can be approximated as $D(\mathrm{L}) p_t = N(\mathrm{L}) \varepsilon_t$, where $D(\mathrm{L}) \equiv 1 - \sum_{k=1}^{p} D_k \mathrm{L}^k$ and $N(\mathrm{L}) \equiv \sum_{k=0}^{q} N_k \mathrm{L}^k$. The polynomials $D(\mathrm{L})$ and $N(\mathrm{L})$ share no common factors. When the approximation is exact, we have the relation $A(\mathrm{L}) = D(\mathrm{L})^{-1} N(\mathrm{L})$ and the equilibrium admits a finite-order ARMA representation. In the frequency domain, ARMA processes are identified as rational functions (i.e., ratios of two polynomial functions).

We offer two geometric interpretations of the connection between the frequency domain's analytic functions and the time domain's impulse response functions. First, suppose the analytic function under consideration is rational. In this case, its geometric properties are completely summarized by the order and location of its poles (roots of the denominator) and zeros (roots of the numerator). The pole-zero plot, which is used extensively in the signal processing literature [e.g., Oppenheim, Willsky and Young (1983)], determines the behavior of impulse responses in the time domain. For example, the causal-stationarity of the impulse responses corresponds to poles located outside the unit circle. Given causal-stationarity, the existence of zeros and

[6] $\mathbf{H}^2(\mathbb{D})$ is a Hilbert space of analytic functions in the open unit disk with square integrability on the boundary. There is a one-to-one mapping between $\mathbf{H}^2(\mathbb{D})$ and the Lebesgue space $\mathbf{L}^2(\mathbb{T})$, where $\mathbb{T}$ denotes the unit circle $\{z \in \mathbb{C} : |z| = 1\}$. Here we use a less well-known version of the Riesz-Fischer theorem to focus on the causal equilibrium with one-sided MA representation. See Rudin (1987) and Lindquist and Picci (2015) for textbook treatments. Isometric isomorphism is defined as bijective mappings that preserve distance.

multiple poles implies non-monotonic and hump-shaped dynamics, and the existence of negative poles implies oscillatory behavior.[7] We summarize these patterns in Figure 1 using three simple examples.

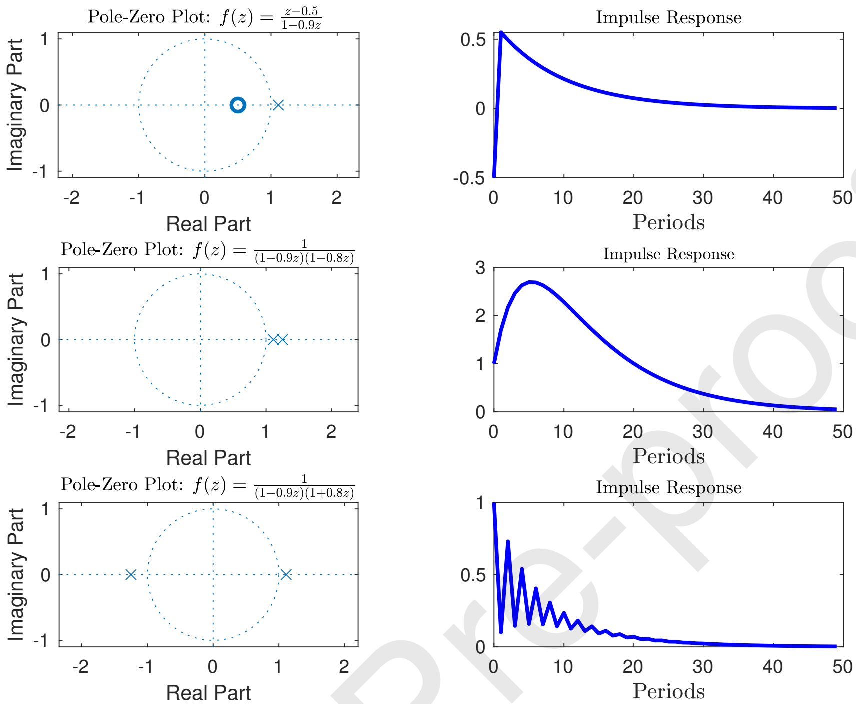

Figure 1: The upper panel plots the case where a zero from MA(1) creates non-monotonic dynamics; the middle panel plots the case where two poles from AR(2) create hump-shaped dynamics; the lower panel plots the case where a negative pole from AR(2) creates oscillatory dynamics.

Second, for a more general class of analytic functions, the sequences of impulse responses $\{A_k\}_{k=0}^{\infty}$ and $\{B_k\}_{k=0}^{\infty}$ correspond to the Laurent series expansions of the analytic functions $A(z)$ and $B(z)$ in the open unit disk $\mathbb{D}$. In this region, the Taylor series coincides with the Laurent series so the impulse responses are linked to the derivatives of analytic functions at the origin ($z = 0$): $A_k = \frac{A^{(k)}(0)}{k!}$, $B_k = \frac{B^{(k)}(0)}{k!}$, where $A^{(k)}(\cdot)$ and $B^{(k)}(\cdot)$ denote the k-th order derivatives. Therefore, the dynamic behavior in the time domain is connected to the local smoothness property of analytic functions around the origin. For example, the non-rational function $\log(1 + \rho z)$ with $|\rho| < 1$ admits the following Taylor series expansion $\log(1 + \rho z) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}\rho^n}{n} z^n$, which produces oscillatory dynamics with alternating signs.

[7]This interpretation shares the same spirit of Rondina and Walker (2018) where endogenous non-invertibility in the signal process creates waves of optimism and pessimism. The resulting dynamics display oscillatory and hump-shaped patterns.

2.3 APFI Approach

The individual expectation $\mathbb{E}_{i,t}p_{t+1}$ is measurable with respect to investor i's signal set. Therefore, we conjecture that the signal representation of the individual expectation is given by $\mathbb{E}_{i,t}p_{t+1} = F_s(\mathrm{L})s_{i,t} + F_p(\mathrm{L})p_t$, where the lag polynomials $F_s(\mathrm{L})$ and $F_p(\mathrm{L})$ are yet to be determined in the equilibrium. Using (2.5) and the signal representation for $\mathbb{E}_{i,t}p_{t+1}$, we can express the equilibrium condition (2.1) as

$$A(\mathrm{L})\varepsilon_t + B(\mathrm{L})\eta_t = \beta\left[F_s(\mathrm{L})\varepsilon_t + F_p(\mathrm{L})p_t\right] + \frac{1}{1-\rho\,\mathrm{L}}\varepsilon_t + \eta_t$$

where the idiosyncratic innovations $\nu_{i,t}$ are washed out by aggregation (2.4). Using the method of undetermined coefficients to match the polynomials associated with each exogenous innovation, we obtain

$$A(z) = \beta\left[F_s\left(A(z), B(z)\right) + F_p\left(A(z), B(z)\right)A(z)\right] + \frac{1}{1-\rho z} \tag{2.7}$$

$$B(z) = \beta F_p\left(A(z), B(z)\right)B(z) + 1 \tag{2.8}$$

where we apply the z-transform to replace the lag operator L with the complex variable z.

(2.7)–(2.8) is a system of functional equations in $\mathbf{H}^2(\mathbb{D})$. $F_s(\cdot,\cdot)$ and $F_p(\cdot,\cdot)$ are operators of analytic functions originated from evaluating conditional expectations in the frequency domain. Since information is endogenous, they are nonlinear in $A(z)$ and $B(z)$. The exact functional forms for $F_s(\cdot,\cdot)$ and $F_p(\cdot,\cdot)$ are derived in Online Appendix S4. As Makarov and Rytchkov (2012) and Huo and Takayama (2018) point out, the underlying system admits no finite-state representation in general and therefore closed-form solutions are not feasible. In other words, the equilibrium values of $A(z)$ and $B(z)$ do not correspond to the z-transform of exact finite-order ARMA processes.

Standard frequency-domain approaches are inherently analytical, which require complicated exploration of the analyticity properties of $A(z)$ and $B(z)$. More importantly, they cannot handle the case when closed-form solutions become unavailable due to information endogeneity. Instead, we advocate a simple iterative approach to solve for the analytic functions $A(z)$ and $B(z)$, which we call analytic policy function iteration (APFI). Motivated by the classic projection methods in the time domain [see, e.g., Judd (1998)], the centerpiece of our approach is the fit of a basis function to the numeric values of $A(z)$ and $B(z)$ over a discretized set of grid points in the complex plane. In this way, we transform the system of nonlinear functional equations (2.7)–(2.8) into a system of linear algebraic equations.

To fix the idea, suppose $\eta_t = 0$, $\forall t$. Then, we project the true solution of $A(z)$ onto the space

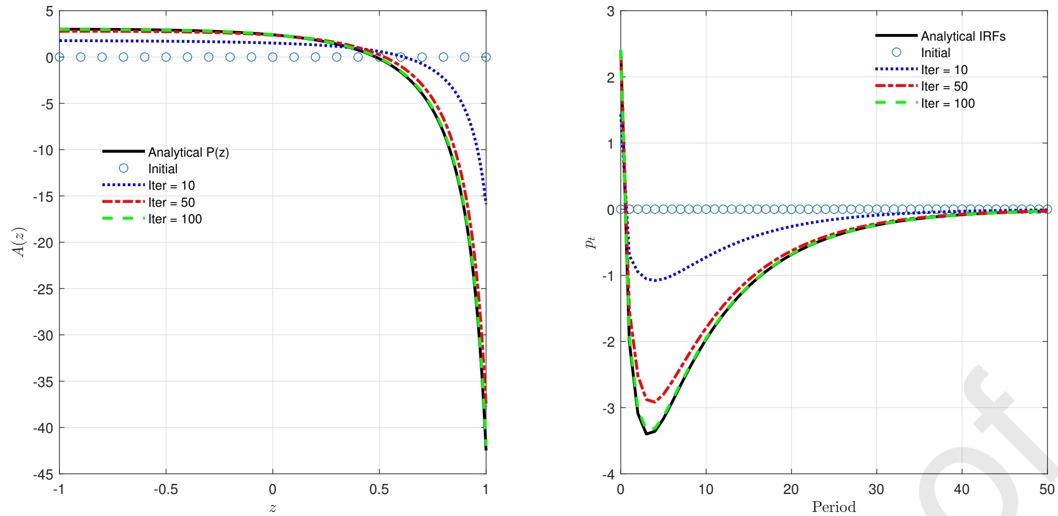

Figure 2: Analytic policy functions $A_k(z)$ (left panel) and implied impulse responses of p_t to $\varepsilon_0 = 1$ (right panel). Parameter values are fixed at: $\beta = 0.98$, $\rho = 0.9$, $\gamma = 1.5$, $\sigma_\varepsilon = 1$, and $\sigma_\nu = 3$.

of rational functions

$$A(z) \approx \frac{N_0 + N_1 z + \cdots + N_q z^q}{1 - D_1 z - \cdots - D_p z^p}, \qquad z \in \mathbb{D} \tag{2.9}$$

with a finite number $(p + q + 1)$ of real-valued coefficients $\{D_1, \ldots, D_p\}$ and $\{N_0, N_1, \ldots, N_q\}$. The choice of rational functions (ARMA processes) as the basis function is natural. The class of ARMA processes serves as the cornerstone for approximating covariance-stationary time series. More importantly, rational functions are vital in numerical evaluations of expectational variables via the Wiener-Hopf optimal prediction formula, as will be shown below.

In our numerical algorithm, we iterate on the function values of $A(z)$ over a properly chosen set of grid points $\{z_j\}_{j=1}^N$ in $\mathbb{D}$. At each iteration, these function values $\{A(z_j)\}_{j=1}^N$ are used to perform the projection. For example, if $A(z)$ is indeed rational, the projection amounts to solving a linear (collocation) equation. In this case, the ARMA coefficients resulting from the projection fully characterize the behavior of $A(z)$ on $\mathbb{D}$. Meanwhile, there is no need to study the analyticity of $A(z)$, which involves heavy algebraic work.[8] We iterate this procedure until the set of function values $\{A(z_j)\}_{j=1}^N$ converges.

To illustrate the accuracy of our method, we consider a special case of the simple model with closed-form solution. In this case, we shut down the transitory shock η_t and model the persistent

[8]In the existing literature, analyticity requires that potential poles (i.e., singular points at which $A(z)$ are not analytic) of $A(z)$ inside the unit circle be removed via lengthy algebraic procedures.

shock as an ARMA$(1,1)$ process $d_t = \rho d_{t-1} + \varepsilon_t - \gamma\varepsilon_{t-1}$ with $0 < \rho < 1$ and $\gamma \neq 0$.[9] Figure 2 compares the exact solution $A(z)$ with the approximated ones $A_k(z)$, where k stands for the number of iterations. The left panel shows $A_k(z)$ quickly converges to the true policy function $A(z)$ in the frequency domain. The right panel plots the convergence of the implied impulse response function in the time domain. In Online Appendix S4 we show the convergence result for the more general case.

2.4 Comparison with Existing Approaches

An essential difference between our approach and the conventional ARMA approximation techniques is that our policy function iteration targets the finite set of function value $\{A(z_j)\}_{j=1}^N$, while the rational function only serves as an intermediate tool to facilitate the computation. Therefore, we do not have to worry about the exact ARMA form in updating the candidate solution.

The symbolic nature of the existing frequency-domain strategies entails heavy algebraic work even for small-scale models (like the above example). In contrast, our projection method is entirely numerical and thus computationally efficient, even for high-dimensional DSGE models. In particular, the policy function iteration features only one state variable (i.e., z) regardless of the model size. Therefore, it does not suffer from the curse of dimensionality.

By truncating the infinite dimension of higher-order expectations (HOEs), Nimark (2017) proposes a state-space approach to compute the approximated equilibrium with endogenous information. Compared to this approach, our APFI method delivers more flexibility. First, a truncated VAR(1) representation of the HOE-augmented state space may become unavailable. This problem appears in many models with endogenous state variables (e.g., capital), non-invertible ARMA shocks (i.e., confounding dynamics), or non-diminishing HOEs (e.g., (2.4) does not hold). Second, when agents' information structures are ex-ante different, the truncation approach becomes infeasible due to the explosion of cross-expectations in the state space. On such occasions, our approach is more suitable. Lastly, the number of unknown parameters involved in the APFI method is significantly smaller than the truncation approach, where the entire transition matrix needs to be solved.

[9] Rondina and Walker (2018) derive the analytical solution as

$$p_t = \underbrace{\frac{1}{\mathrm{L}-\beta}\left[\mathrm{L}\,D(\mathrm{L}) - \beta D(\beta)\frac{h(\mathrm{L})}{h(\beta)}\right]}_{A(\mathrm{L})}\varepsilon_t, \quad D(\mathrm{L}) = \frac{1-\gamma\,\mathrm{L}}{1-\rho L}, \quad h(\mathrm{L}) = \psi\theta + (1-\psi)\frac{\theta - \mathrm{L}}{1-\theta\,\mathrm{L}}$$

where $\theta \in (-1,1)$ and β are the only two distinct solutions inside the unit circle of the equation $zD(z)h(\beta) - \beta D(\beta)h(z) = 0$ and $\psi = \sigma_\varepsilon^2/(\sigma_\nu^2 + \sigma_\varepsilon^2)$. It can be shown when such a θ exists, p_t follows an ARMA$(2,2)$ process satisfying $A(\theta) = 0$.

3 Theoretical Framework

This section establishes the theoretical foundation of our method. Section 3.1 constructs a canonical form for a general class of dynamic incomplete-information models, which will be used in our baseline numerical algorithm. In Section 3.2, we provide three theorems that characterize the key ingredients in the APFI framework. The appendix contains the proof of the theorems in this section.

3.1 Canonical Representation

We study a class of linear or linearized rational expectations models with general information structures. We cast the set of model equilibrium conditions into the following system of n_x linear expectational difference equations

$$\sum_{k=0}^{l} A_k y_{t-k} + \sum_{k=0}^{h} B_k \mathbb{E}_t y_{t+k} = \mathbf{0}_{n_x \times 1} \tag{3.1}$$

where the model variables y_t and their coefficient matrices (A_k, B_k) are partitioned as

$$y_t \equiv \begin{bmatrix} x_t \\ a_t \\ s_t \end{bmatrix}, \qquad A_k \equiv \begin{bmatrix} A_k^x & A_k^a & A_k^s \end{bmatrix}, \qquad B_k \equiv \begin{bmatrix} B_k^x & B_k^a & B_k^s \end{bmatrix}$$

and $\mathbf{0}_{n_x \times 1}$ is a $n_x \times 1$ vector of zeros. The model system (3.1) comprises a non-expectational block (i.e., "A" block) and an expectational block (i.e., "B" block).[10] $\mathbb{E}_t(\cdot)$ is a generalized mathematical expectation operator conditional on certain information sets at time t, as will be explained below. Our formulation adopts the timing convention that a variable is dated t if it is realized at t, so there is no need to specify which elements of y_t are predetermined; the structure of the coefficient matrices $\{A_k\}$ automatically pins down the list of predetermined variables. For example, if the current capital stock k_t is predetermined (i.e., realized at $t-1$), then it will be treated as a $t-1$ variable.

There are three types of model variables. First, s_t is a $n_s \times 1$ vector of exogenous shocks with $n_x \times n_s$ coefficient matrices $\{A_k^s, B_k^s\}$. It follows a covariance-stationary VARMA(p_s, q_s) process

[10] It is necessary to include both the A_0 and B_0 coefficient matrices in (3.1) because the nowcast $\mathbb{E}_t y_t$ does not necessarily equal its realization y_t under certain type of information structure. For such an example, see the dispersed information version of the Phillips curve in Online Appendix S5. Another reason is that some elements of y_t pertain to agents' own decisions and hence are measurable with respect to their information sets. In the computer program, however, these variables are not explicitly included in the agents' signal sets used in computing expectations. Therefore, the user needs to associate them only with the A_0 coefficient matrix (i.e., outside the expectation operator) to avoid any measurability issue caused by the computer.

driven by exogenous innovations

$$s_t = \sum_{k=1}^{p_s} C_k^s s_{t-k} + \sum_{k=0}^{q_s} D_k^s \epsilon_{t-k}, \qquad \epsilon_t \sim \mathbb{N}(\mathbf{0}_{n_\epsilon \times 1}, \Sigma_\epsilon) \tag{3.2}$$

where $\{C_k^s, D_k^s\}$ are $n_s \times n_s$ and $n_s \times n_\epsilon$ coefficient matrices, respectively.[11] ϵ_t is a $n_\epsilon \times 1$ vector of *i.i.d.* Gaussian innovations with positive definite covariance matrix $\Sigma_\epsilon \succ 0$. It contains both individual and aggregate innovations, which are allowed to be arbitrarily correlated. In particular, the individual innovations satisfy the law of large numbers. Second, x_t is a $n_x \times 1$ vector of endogenous variables with $n_x \times n_x$ coefficient matrices $\{A_k^x, B_k^x\}$. It contains both individual choices and aggregate outcomes. While individual choices in x_t may depend on all innovations in ϵ_t, aggregate outcomes in x_t only respond to aggregate innovations in ϵ_t. Lastly, $a_t = \int_0^1 x_t di$ is the aggregation of x_t across a continuum of agents indexed by $i \in [0, 1]$ with $n_x \times n_x$ coefficient matrices $\{A_k^a, B_k^a\}$. Specifically, a_t corresponds to the component of x_t with respect to the aggregate innovations. As an illustrative example, Online Appendix S4 shows how to cast the simple model of Section 2 into the canonical form (3.1)–(3.2).

3.1.1 Expectational Block and Information Structure

The canonical form (3.1)–(3.2) allows for a flexible specification of the model environment and its information structure. The model economy supported by the system (3.1) consists of a continuum of agents with constant measure. Each agent is subject to three types of shocks: (i) economy-wide aggregate shocks that affect all agents; (ii) group-specific shocks that only affect a particular group or type of agents with non-zero measure; and (iii) idiosyncratic shocks that satisfy the law of large numbers. While the first two types of shocks affect aggregate dynamics, the third type washes out upon aggregation.

The generalized expectation operator $\mathbb{E}_t(\cdot)$ that appears in (3.1) allows for heterogeneous conditional expectations. That is, every time a variable in y_t appears in the expectational block of (3.1), its expectation is allowed to be associated with a different information set. To understand this feature more precisely, suppose there are M types of expectations involved in the system. Then (3.1) can be expanded as

$$\sum_{k=0}^{l} A_k y_{t-k} + \sum_{m=1}^{M} \sum_{k=0}^{h} B_{m,k} \mathbb{E}_{m,t} y_{t+k} = \mathbf{0}_{n_x \times 1} \tag{3.3}$$

where we partition the original "B" coefficient matrices in (3.1) into a set of sparse matrices.[12]

[11] The VARMA(p_s, q_s) process (3.2) is covariance-stationary if and only if $\det(\mathrm{I}_{n_s} - \sum_{k=1}^{p_s} C_k^s z^k) \neq 0$ for all $z \in \mathbb{D}$, where det denotes the determinant operator and I_{n_s} is a $n_s \times n_s$ identity matrix.

[12] We thank an anonymous referee for suggesting this compact representation.

The types of expectations are embedded in the index $m = 1, 2, \ldots, M$, including individual expectations with respect to different information sets as well as group expectations when information is homogeneous within each group but heterogeneous across groups. The information set for each $\mathbb{E}_{m,t}$, denoted by $\Omega_{m,t}$, consists of the smallest closed subspace generated by the history of *exogenous* signals $s^t_{m,\Omega} \equiv \{s_{m,\Omega,t}, s_{m,\Omega,t-1}, \ldots\}$ and that of *endogenous* signals $x^t_{m,\Omega} \equiv \{x_{m,\Omega,t}, x_{m,\Omega,t-1}, \ldots\}$, denoted by $\Omega_{m,t} = s^t_{m,\Omega} \vee x^t_{m,\Omega}$. These two types of signals are defined as subsets of exogenous shocks and endogenous variables, respectively, i.e., $s_{m,\Omega,t} \subseteq s_t$ and $x_{m,\Omega,t} \subseteq x_t$.

In addition, $\mathbb{E}_{m,t}$ in (3.3) may refer to composite expectation operators defined as linear combinations of individual expectations about a variable in y_t. Some examples include (i) economy-wide average expectations, (ii) group or type-specific average expectations, and (iii) finite combinations of individual expectations. For simplicity, in the discussion that follows, we focus on the first two types of compositions.

In our numerical toolbox, we adopt the reductive representation (3.1) to lower the user's input cost. Instead of supplying a set of sparse matrices, the user will simply specify the information set for each equation in the system that involves expectations. Online Appendix S3 provides a user guide on this specification as well as the use of dummy variables in handling the case of multiple types of expectations in the same equation.

Altogether, our canonical form can accommodate models with four different types of information structure. The first type (termed as I.0) refers to the "imperfect, homogeneous information" in which agents are equipped with the same signal set Ω_t. The second type (termed as I.1) refers to the "dispersed information", where the signal structures are ex-ante identical, but idiosyncratic (mean zero) realizations imply that individual-level expectations are ex-post different and all noises aggregate to zero. The third type (termed as I.2) considers the "asymmetric information" in which the signal structures are not ex-ante identical, but expectations are common within groups. This type of information structure precludes purely idiosyncratic noises so that $\Omega_{i,m,t} = \Omega_{m,t}$ for any agent i in group m. The fourth type (termed as I.3) considers the "asymmetric, dispersed information", where the signal sets are different across groups, and idiosyncratic (mean-zero) realizations cause dispersion within groups.

Table 1 summarizes the types of information structure along with the corresponding examples from the literature. While the time-domain truncation algorithm of Nimark (2017) focuses primarily on the dispersed information setup (I.1), our framework has broader applicability and covers most models in the literature.[13]

[13] The limitation of Nimark (2017) lies in the explosion of cross-expectations in the state space whenever information is "asymmetric" as defined above.

Information Structure	Literature Example
I.0 (imperfect, homogeneous)	Blanchard, L'Huillier and Lorenzoni (2013); Chahrour and Jurado (2018)
I.1 (dispersed)	Graham and Wright (2010); Nimark (2008, 2017); Melosi (2017); Adams (2021)
I.2 (asymmetric)	Barsky and Sims (2012); Kasa, Walker and Whiteman (2014); Tang (2015); Kohlhas (2019)
I.3 (asymmetric, dispersed)	HANK model of Angeletos and Huo (2021) (see Section 5.2)

Table 1: Summary of information structures. Nimark (2008) and Melosi (2017) consider two types/groups of agents: households and firms. Households are equipped with full information while firms are subject to dispersed information. The aggregation of expectations involves only group average among firms.

3.1.2 Frequency-Domain Equilibrium

Now we derive the equilibrium fixed point of the model system (3.1) in the frequency domain. For clarity of exhibition, we assume the expectational block is equipped with only one information set Ω_t (i.e., I.0). All results derived herein can be easily generalized to more complicated information structures I.1–I.3.

We work with the VMA(∞) representation of the model variables

$$y_t = \begin{bmatrix} x_t \\ a_t \\ s_t \end{bmatrix} = \begin{bmatrix} \Gamma^x(\mathrm{L}) \\ \Gamma^a(\mathrm{L}) \\ \Gamma^s(\mathrm{L}) \end{bmatrix} \epsilon_t \equiv \Gamma^y(\mathrm{L})\epsilon_t, \qquad \Gamma^y(\mathrm{L}) = \sum_{k=1}^{\infty} \Gamma^y_k \, \mathrm{L}^k \tag{3.4}$$

in the vector space spanned by the history of structural innovations $\epsilon^t \equiv \{\epsilon_t, \epsilon_{t-1}, \ldots\}$. Some remarks about (3.4) are in order. First, the equilibrium solution is defined as $x_t = \Gamma^x(\mathrm{L})\epsilon_t$, which is a matrix generalization of (2.5). In particular, $\Gamma^x(\mathrm{L})$ is a $n_x \times n_\epsilon$ matrix of lag polynomials with the (i, j)-th element in Γ^x_k measuring exactly the impulse response of the i-th variable in x_{t+k} to the j-th innovation in ϵ_t. Second, given the solution $\Gamma^x(\mathrm{L})$, its aggregation $\Gamma^a(\mathrm{L})$ can be obtained by nullifying those columns of $\Gamma^x(\mathrm{L})$ corresponding to the idiosyncratic components in ϵ_t. Third, the shock representation $\Gamma^s(\mathrm{L})$ can be computed directly from its VARMA form (3.2).

As in Section 2, we implement the z-transform of $\Gamma^x(\mathrm{L})$ to obtain its functional equivalent in the frequency domain $\Gamma^x(z) \in \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$. $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ denotes the generalization of the univariate Hardy space for $n_x \times n_\epsilon$ matrix of analytic functions, with each element $\Gamma^x_{(i,j)}(z) \in \mathbf{H}^2(\mathbb{D})$.[14] The

[14]Such a multivariate generalization is trivial using the Hilbert-Schmidt (trace) norm. See Lindquist and Picci (2015).

second moment properties of $\{x_t\}$ are fully characterized by its spectral density function $S_x(\cdot)$, defined as

$$S_x(\omega) = \frac{1}{2\pi}\Gamma^x\left(e^{-i\omega}\right)\Sigma_\epsilon\Gamma^x\left(e^{i\omega}\right)', \qquad \omega \in [-\pi, \pi] \tag{3.5}$$

where ω denotes the frequency and $i^2 = -1$. (3.5) is a special case of the covariance-generating function evaluated on the unit circle (with normalization). Similarly, we can obtain the z-transforms $\Gamma^a(z) \in \mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$ and $\Gamma^s(z) \in \mathbf{H}^2_{n_s\times n_\epsilon}(\mathbb{D})$. Collectively, it follows that $\Gamma^y(z) = [\Gamma^x(z)', \Gamma^a(z)', \Gamma^s(z)']' \in \mathbf{H}^2_{(2n_x+n_s)\times n_\epsilon}(\mathbb{D})$.

We collect the endogenous and exogenous signals to derive the information set as

$$\Omega_t = \begin{bmatrix} x_{\Omega,t} \\ s_{\Omega,t} \end{bmatrix} = \begin{bmatrix} \Gamma^{x\Omega}(\mathrm{L}) \\ \Gamma^{s\Omega}(\mathrm{L}) \end{bmatrix}\epsilon_t \equiv \Gamma^{\Omega}(\mathrm{L})\epsilon_t \tag{3.6}$$

where $\Gamma^{x\Omega}(\mathrm{L})$ is a sub-block matrix of $\Gamma^x(\mathrm{L})$ corresponding to the endogenous signals, $\Gamma^{s\Omega}(\mathrm{L})$ is a sub-block matrix of $\Gamma^s(\mathrm{L})$ corresponding to the exogenous signals, and $\Gamma^{\Omega}(\mathrm{L})$ is the VMA(∞) representation of the information set. We then compute the conditional expectations in the frequency domain using the celebrated Wiener-Hopf optimal prediction formula

$$\mathbb{E}[y_{t+k}|\Omega^t] = \underbrace{\left[\mathrm{L}^{-k}\,\Gamma^y(\mathrm{L})\Sigma_\epsilon\Gamma^\Omega\left(\mathrm{L}^{-1}\right)'\left(\widetilde{\Gamma}^\Omega\left(\mathrm{L}^{-1}\right)'\right)^{-1}\right]_+\Sigma_u^{-1}\widetilde{\Gamma}^\Omega(\mathrm{L})^{-1}}_{F_k(\Gamma^y(\mathrm{L}))}\Omega_t, \qquad \forall k \tag{3.7}$$

where the annihilation operator $[\cdot]_+$ removes the negative-power part of the function expressed in terms of its series expansion, and $F_k(\Gamma^y(\mathrm{L}))$ is the signal VMA(∞) representation of $\mathbb{E}[y_{t+k}|\Omega^t]$. The information set admits the Wold fundamental representation $\Omega_t = \widetilde{\Gamma}^\Omega(\mathrm{L})u_t$, where u_t is a vector $i.i.d.$ innovation process with covariance matrix Σ_u.[15] In Online Appendix S2, we provide computational methods for finding the analytic function $\widetilde{\Gamma}^\Omega(z)$ via factorization techniques of (3.5).

The annihilation operator $[\cdot]_+$ is a linear operator in the space of analytic functions. Therefore, it is straightforward to generalize (3.7) to the expectation of discounted future sum

$$\sum_{k=0}^{\infty}\beta^k\mathbb{E}[y_{t+k}|\Omega^t] = \underbrace{\left[\frac{\mathrm{L}}{\mathrm{L}-\beta}\Gamma^y(\mathrm{L})\Sigma_\epsilon\Gamma^\Omega\left(\mathrm{L}^{-1}\right)'\left(\widetilde{\Gamma}^\Omega\left(\mathrm{L}^{-1}\right)'\right)^{-1}\right]_+\Sigma_u^{-1}\widetilde{\Gamma}^\Omega(\mathrm{L})^{-1}}_{F_k(\Gamma^y(\mathrm{L}),\,\beta)}\Omega_t, \tag{3.8}$$

where $\beta \in (0,1)$ is the discount rate. Formula (3.8) is useful in many economic applications. For example, Section 5.2 illustrates how our numerical toolbox can handle the type of expectations

[15] We can always orthogonalize u_t by performing eigen-decomposition of Σ_u.

in (3.8).

Substituting (3.4), (3.6), and (3.7) into (3.1) and applying the z-transform yield a system of functional equations in the unknown policy functions $\Gamma^y(z)$

$$\sum_{k=0}^{l} A_k z^k \Gamma^y(z) + \sum_{k=0}^{h} B_k F_k(\Gamma^y(z))\Gamma^{\Omega}(z) = 0, \qquad z \in \mathbb{D} \tag{3.9}$$

(3.9) is a restatement of the equilibrium condition in the space of analytic functions $\mathbf{H}^2_{(2n_x+n_s)\times n_\epsilon}(\mathbb{D})$, from which we derive a fixed-point condition used in our iterative algorithm.

3.2 Foundation

We solve the functional equations (3.9) using a projection method that fits the function values $\Gamma^x(z_j)$ over a discretized set of grid points in the open unit disk, $z_j \in \mathbb{D}$, $j = 1, 2, \ldots, N$. Like other global solution methods, the projection performance hinges on the choice of an appropriate basis function. In our iterative framework, we approximate $\Gamma^x(z)$ using a VARMA(p_x, q_x) process

$$x_t \approx \sum_{k=1}^{p_x} C_k^x x_{t-k} + \sum_{k=0}^{q_x} D_k^x \epsilon_{t-k} \tag{3.10}$$

or equivalently, a rational function in the frequency domain

$$\Gamma^x(z) \approx C^x(z)^{-1} D^x(z), \qquad C^x(z) \equiv \mathrm{I}_{n_x} - \sum_{k=1}^{p_x} C_k^x z^k, \qquad D^x(z) \equiv \sum_{k=0}^{q_x} D_k^x z^k \tag{3.11}$$

where $\{C_k^x, D_k^x\}$ are $n_x \times n_x$ and $n_x \times n_\epsilon$ coefficient matrices, respectively. When the objective function $\Gamma^x(z)$ is rational, e.g., in the exogenous information case, the approximation is exact in theory.

The classical idea of approximating analytic functions using rational functions dates back to the Runge theorem [see, e.g., Rudin (1987)]. However, our objective here is more involved as we are solving for analytic functions in $\mathbf{H}^2(\mathbb{D})$ that correspond to sequences of square-summable impulse responses. Thus, the approximation accuracy needs to be established under the appropriate norm related to covariance-stationarity. To this end, we establish the following theorem on the denseness condition of rational functions in $\mathbf{H}^2(\mathbb{D})$ that justifies our approximation.

Theorem 3.1 (Denseness of Rational Functions)**.** *Define the set of $n_x \times n_\epsilon$ matrices of rational analytic functions that correspond to VARMA(p, q) processes as $\mathbf{Q}_{(p,q)}$, where each element of*

$\mathbf{Q}_{(p,q)}$ *is of the form*

$$\mathbf{Q}_{(p,q)}^{(m,n)} := \left\{ c^{(m,n)} \frac{\prod_{j=1}^{q}(1-b_j^{(m,n)}z)}{\prod_{i=1}^{p}(1-a_i^{(m,n)}z)} : \ a_i^{(m,n)}, b_j^{(m,n)}, c^{(m,n)} \in \mathbb{C}, \ |a_i^{(m,n)}| < 1, \ \forall i,j \right\}$$

for $m = 1, 2, \ldots, n_x$ *and* $n = 1, 2, \ldots, n_\epsilon$*. Then* $\bigcup_{p,q\,\in\mathbb{N}} \mathbf{Q}_{(p,q)}$ *is dense in* $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$*.*

It follows from Theorem 3.1 that the set of rational analytic functions is dense in $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$. The key implication of this result is that one can always use a set of VARMA(p, q) processes to replicate any covariance-stationary equilibrium dynamics. In the mean squared sense, the approximation can achieve arbitrary accuracy, provided that the underlying (true) equilibrium process is not too persistent or too close to having a unit root.

Next, we construct the appropriate state-space grid of points $z_j \in \mathbb{D}$, $j = 1, 2, \ldots, N$, in the open unit disk. We choose the grid points within the real subset $\mathbb{U} = (-1, 1) \subset \mathbb{D}$. By placing $\{z_j\}_{j=1}^N$ in the unit interval, we bypass the significant complication of handling complex numbers in the numerical algorithm that could compromise accuracy. However, two questions naturally arise from this restriction. First, is a solution to the functional equation (3.9) restricted to $\mathbb{U}$ a legitimate covariance-stationary equilibrium of the model? Second, does the projection of the function values of $\Gamma^x(z)$ defined on $\mathbb{U}$ deliver an accurate approximation to the true function, which is defined on the bigger set $\mathbb{D}$? To address these questions, we establish the following theorem using the theory of analytical continuation and the theory of convergence of analytic functions.

Theorem 3.2 (Analytic Continuation and Convergence Criterion)**.** *Suppose* $\Psi^x(z) \in \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ *and let* $\Psi^y(z) = [\Psi^x(z)', \Psi^a(z)', \Gamma^s(z)']'$ *satisfy*

$$\sum_{k=0}^{l} A_k z^k \Psi^y(z) + \sum_{k=0}^{h} B_k F_k(\Psi(z))\Psi^\Omega(z) = 0, \qquad z \in \mathbb{U} = (-1, 1) \tag{3.12}$$

with $\Psi^\Omega(z) = [\Psi^{x\Omega}(z)', \Gamma^{s\Omega}(z)']'$ *for a given selection of endogenous and exogenous signals. Then a stationary equilibrium exists for model* (3.1) *and is given by* $\Gamma^x(z) = \Psi^x(z)$*. More generally, let* $\Gamma^x(z)$ *be any analytic function in the Hardy space* $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$*. If there exists an analytic function* $\Psi^x(z)$ *such that* $\Psi^x(z) = \Gamma^x(z)$ *for all* $z \in (-1, 1)$*, then* $\Gamma^x = \Psi^x$ *on the entire open unit disk* $\mathbb{D}$*. Finally, if a sequence of rational functions* $\{\Gamma^x_n(z)\}_{n\in\mathbb{N}} \in \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ *converges pointwise in* $\mathbb{U}$ *to a function* $\Gamma^x(z)$*, then* $\lim_{n\to\infty} \|\Gamma^x(z) - \Gamma^x_n(z)\|_{\mathbf{H}^2} = 0$*.*

The first part of Theorem 3.2 states that the analytic continuation of the solution to (3.12) is the solution to the analytic continuation of the functional equation (3.12), which is (3.9). Therefore, solving the equation over the grid from $\mathbb{U}$ is sufficient to deliver the model equilibrium. This result is analogous to the law of permanence of functional equations. By the uniqueness of the analytic continuation, an immediate implication is that if (3.12) has a unique solution, the model admits a unique equilibrium. In our numerical algorithm, the policy function values within each iteration emerge from rational functions. By the second part of Theorem 3.2, the projection based on $\mathbb{U}$ delivers an accurate reconstruction of the original policy function defined on $\mathbb{D}$. More importantly, when iterations on the finite vector $\{\Gamma^x(z_j)\}_{j=1}^N$ converge in the pointwise manner, we find the fixed point of the functional equations in $\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$. This result is ensured by the last part of Theorem 3.2, which establishes a convergence criterion for our numerical algorithm.

In our APFI framwork, we compute expectational variables using the Wiener-Hopf optimal prediction formula, which requires computing the annihilation operator $[\cdot]_+$ for the function

$$\Theta(z) \equiv z^{-k}\Gamma^x(z)\Sigma_\epsilon\Gamma^\Omega\left(z^{-1}\right)'\left(\widetilde{\Gamma}^\Omega\left(z^{-1}\right)'\right)^{-1} \tag{3.13}$$

Without loss of generality, we assume now that $\Theta(z)$ is univariate but all analyses contained here continue to apply to the case of matrix-valued function. By inspection, $\Theta(z)$ is not analytic in the open unit disk $\mathbb{D}$ due to the cross-spectral density term. Future expectations also induce poles at $z = 0$. Elementary complex analysis indicates that $\Theta(z)$ has different power (Laurent) series expansions in different regions of convergence (ROC) inside $\mathbb{D}$. Different ROCs are partitioned as annuli centered at the origin ($z = 0$) according to the positions of singularities (poles) on the complex plane.[16] Given an annulus $R_1 < |z| < R_2$ with $0 < |R_1| < |R_2| < 1$, we can define its two-sided Laurent series expansion and the corresponding annihilation as

$$\Theta(z) = \sum_{k=-\infty}^{\infty}\Theta_k z^k, \qquad [\Theta(z)]_+ = \sum_{k=0}^{\infty}\Theta_k z^k$$

Then the questions we face are: what is the appropriate region to perform the annihilation in computing the Wiener-Hopf prediction, and how do we calculate the annihilation? The following theorem provides the answers to these questions.

Theorem 3.3 (Annihilation)**.** *Suppose $\Gamma^x(z)$ is rational and the Wold fundamental function (spectral factor) $\widetilde{\Gamma}^\Omega(z)$ is invertible on the closed unit disk, i.e., $\widetilde{\Gamma}^\Omega(z)^{-1}$ is analytic on the closed unit disk.*[17] *Then the coefficient matrices $\{\Theta_k\}_{k=0}^\infty$ in the annihilation of $\Theta(z)$ in (3.13) are given by the inverse discrete time Fourier transform (IDTFT) around the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| =$*

[16] The open unit disk $\mathbb{D}$ is a special annulus $0 \leqslant |z| < 1$ with the exception that $z = 0$ is in the region of analyticity. One can also define the annulus of ROC that does not center at 0; however, in any particular ROC in which $\Theta(z)$ is analytic, its Laurent series expansion is unique.

[17] The imposition of invertibility is slightly stronger than the Wold fundamentality as analyticity is required on the unit circle. See Forni, Gambetti and Sala (2019).

1}

$$\Theta_k = \left[\frac{1}{2\pi i} \oint_{\mathbb{T}} \Theta(z) z^{-k} \frac{dz}{z} \right], \qquad k = 0, 1, 2, \ldots \tag{3.14}$$

where $\oint$ denotes the (counterclockwise) contour integral. For any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\Theta_N(z) = \sum_{k=-N/2}^{N/2-1} \Theta_k z^k, \qquad \|\Theta_N(z) - \Theta(z)\|_{L^2} < \epsilon$$

where $\|\cdot\|_{L^2}$ denotes the L^2 norm on $L^2(\mathbb{T})$. The positive part of the coefficient matrices $\{\Theta_k\}_{k=0}^{N/2-1}$ can be approximated via the inverse discrete Fourier transform (IDFT)

$$\Theta_k \approx \frac{1}{N} \sum_{n=0}^{N-1} \Theta\left(\exp\left(-i\frac{2\pi n}{N}\right)\right) \exp\left(i\frac{2\pi n}{N}k\right), \qquad k = 0, 1, \ldots, N/2 - 1 \tag{3.15}$$

We treat the signal as exogenous with a rational function representation during each algorithm iteration, even when information is endogenous. In this case, it is easy to show the resulting solution $\Gamma^x(z)$ is also rational, and the first part of Theorem 3.3 implies that the annihilation should be computed in the ROC that includes the unit circle $\mathbb{T}$. This result is connected to the derivation of the Wiener-Hopf formula. Therefore, we compute the annihilation coefficient matrices using the IDTFT formula (3.14).

The second part of Theorem 3.3 considers the numerical approximation of the annihilated function $[\Theta(z)]_+$, which is an infinite series. To this end, we first approximate the original analytic function on the unit circle using the two-sided finite partial sum $\Theta_N(z)$. The approximation can be made arbitrarily accurate with increasing order N by the Riesz-Fischer theorem. We then compute the positive part of this partial sum using the IDFT formula (3.15). IDFT is a classical method of computing the inverse Fourier transform with superior numerical efficiency. In particular, we have used the fact that $\Theta(z)$ can be evaluated at evenly-spaced points on the unit circle indexed by different frequencies $\omega_n = 2\pi n/N$, $n = 0, 1, \ldots, N-1$, using formula (3.13).

The standard way of computing the annihilated function $[\Theta(z)]_+$ in the literature follows Hansen and Sargent (1980). This method characterizes the annihilated function as the difference between the original function and the principal part of its Laurent series expansion around the singularities inside the unit circle. It provides the closed-form expression for the annihilation. Using the residue theorem from complex analysis, one can compute the coefficients associated with the principal part. Since the residue theorem formula uses the values of endogenous functions at the singularities, these functional constants are yet to be determined in the equilibrium. The solution of these intertwined constants then determines the equilibrium existence and uniqueness. Tan and Walker (2015) and Huo and Takayama (2018) generalize this formula to the multivariate case with higher-order (repeated) singularities.

The existing approach incurs a significant amount of symbolic algebra, including lengthy partial fraction simplifications. The equilibrium determination procedure ("roots counting") is also non-trivial [Tan and Walker (2015)]. Such a procedure is infeasible in models with endogenous information since the pole structure of $\Theta(z)$ is not designated ex-ante but endogenous and changes across iterations. In medium or large-scale models, such a method also leads to an algebraic nightmare. In contrast, the discrete Fourier transform method we develop here addresses the underlying problem by sidestepping the functional approach and the equilibrium determination procedure; it uses the IDFT technique to achieve fast and algorithmic computation. This method also connects well with our overall APFI strategy—transforming functional operations into numerical evaluations.

We emphasize that there is one restriction in Theorem 3.3: the forecasting objective cannot have any unit root (i.e., display any random walk property); otherwise, the Riesz-Fischer theorem fails, and the computation is theoretically invalid. Therefore, our baseline APFI algorithm only handles stationary equilibrium systems.

4 Baseline Algorithm

Based on the theoretical framework, we propose the baseline APFI algorithm as follows. According to Theorem 3.1, we begin with a conjecture about the as-yet-unknown matrix of rational functions $\Gamma^x(\cdot)$ in (3.11). This initial set of policy functions for x_t is then used along with Theorem 3.3 to calculate the expectational variables $\{\mathbb{E}_t x_{t+k}\}_{k=0}^h$ in (3.1). Next, we obtain an updated set of policy functions for x_t by solving the functional equations (3.9). By Theorem 3.2, it suffices to solve for the values of $\Gamma^x(z)$ over a discretized set of grid points on the open unit interval. Evaluating (3.9) at these nodes transforms the functional equations into simpler systems of linear algebraic equations. If the distance between the guess and updated policy values is less than a pre-specified criterion, the policy functions have converged to the equilibrium according to Theorem 3.2. Otherwise, we set the updated policy functions as a new guess and repeat the iterations until convergence.

4.1 Implementation

Without loss of generality, we make two simplifying assumptions about our canonical form. These assumptions ease the exposition but do not affect any of the results presented in this section.

First, we eliminate the aggregation variables a_t whose properties are completely summarized by the endogenous variables x_t. From (3.9) we expand the functional fixed-point condition as

$$\sum_{k=0}^{l} A_k^x z^k \Gamma^x(z) + \sum_{k=0}^{l} A_k^s z^k \Gamma^s(z) + \sum_{k=0}^{h} B_k F_k(\Gamma^y(z))\Gamma^{\Omega}(z) = 0, \qquad z \in \mathbb{D} \tag{4.1}$$

where $\Gamma^y(z) = [\Gamma^x(z)', \Gamma^s(z)']'$. Second, we do not distinguish between individual and aggregate variables in equilibrium so that a priori each element in x_t is allowed to depend on all elements in ϵ_t.

The baseline APFI algorithm operates under the following regularity condition.

Assumption 4.1 (Regularity)**.** *The full-information version of the simplified model system* (4.1) *admits a covariance-stationary solution. Let* $A^x(z) = \sum_{k=0}^{l} A_k^x z^k$*. The finite-order polynomial matrix function* $A^x(z)$ *is invertible on the closed unit disk* $\mathbb{D} \bigcup \mathbb{T}$*, i.e., its determinant polynomial* $\det A^x(z)$ *has no root inside the closed unit circle. If* $l = 1$*, the invertibility condition requires that the matrix* $(A_0^x)^{-1} A_1^x$ *has spectral radius smaller than 1, i.e., the absolute value of its eigenvalues are all smaller than unity.*[18]

The stationarity assumption on the full-information solution rules out non-stationarity originating from the model's primitive structure. This restriction is necessary since our baseline APFI algorithm, which is based on the canonical form, cannot handle non-stationary solutions such as random walk equilibria (see the discussion in Section 3.2). Suppose the model admits a stationary solution under full information. In this case, its solution under incomplete information is also likely to be stationary since information frictions generally lead to dampened and sluggish dynamics compared to the full-information case. The regularity condition in Assumption 4.1 holds for a wide range of models, including those with endogenous physical states (see Section 5.1 and 5.2). When $l = 0$ (i.e., no predetermined variables), the invertibility assumption is automatically satisfied.

Assumption 4.1 ensures that our baseline APFI algorithm induces a well-defined (nonlinear) operator $\mathcal{A} : \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D}) \mapsto \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$, and (4.1) becomes the fixed point of the algorithm operator

$$\Gamma^x(z) = \mathcal{A}\left(\Gamma^x(z)\right) \equiv -A^x(z)^{-1} A^s(z) \Gamma^s(z) - A^x(z)^{-1} \left(\sum_{k=0}^{h} B_k F_k(\Gamma^y(z)) \Gamma^{\Omega}(z) \right) \tag{4.2}$$

where $A^s(z) = \sum_{k=0}^{l} A_k^s z^k$. In Proposition 2 of the appendix, we show that with additional restrictions the operator $\mathcal{A}$ is stable (i.e., non-explosive) everywhere. We now describe the

[18] If A_0^x is singular, we compute the pseudo inverse $(A_0^x)^+$ instead. To compute $(A_0^x)^+$, let the singular value decomposition of A_0^x be given by $A_0^x = USV'$, where the matrices U and V satisfy $U'U = V'V = \mathrm{I}$ but are not necessarily square, and S is a square and diagonal matrix with non-zero entries. Then $(A_0^x)^+ = VS^{-1}U'$.

algorithm implementation as follows.

Algorithm 4.2 (Baseline APFI)**.** Given Assumption 4.1, the baseline APFI algorithm is summarized by a sequence of easily implementable steps as follows:

1. **Initialization.** Discretize the state space $\mathbb{U} = (-1, 1)$ into N grid points $\{z_j\}_{j=1}^N$. For simplicity, we let $\{z_j\}_{j=1}^N$ be evenly spaced on $\mathbb{U}$. Set the initial policy function values $\{\Gamma^x(z_j)\}_{j=1}^N$.

2. **Projection.** Fit the z-transform of the VARMA(p_x, q_x) representation for x_t in (3.11) to the set of data points $\{(z_j, \Gamma^x(z_j))\}_{j=1}^N$, i.e.,
$$C^x(z_j)^{-1} D^x(z_j) = \Gamma^x(z_j), \qquad j = 1, 2, \ldots, N \tag{4.3}$$
and obtain the VARMA coefficient matrices $\left\{C_1^x, \ldots, C_{p_x}^x\right\}$ and $\left\{D_0^x, D_1^x, \ldots, D_{q_x}^x\right\}$.

3. **Evaluation.** For each node z_j, $j = 1, 2, \ldots, N$, evaluate the z-transforms of expectational variables $\{F_k(\Gamma^y(z_j))\Gamma^\Omega(z_j)\}_{k=0}^h$ in (4.1) based on the information set (3.6), the Wiener-Hopf formula (3.7), and the fitted policy functions $\Gamma^x(z) = C^x(z)^{-1} D^x(z)$ in step 2.

4. **Updating.** For each node z_j, $j = 1, 2, \ldots, N$, compute the updated policy function value $\hat{\Gamma}^x(z_j)$ implied by (4.1) via solving the following systems of linear algebraic equations
$$A^x(z_j)\hat{\Gamma}^x(z_j) = -A^s(z_j)\Gamma^s(z_j) - \sum_{k=0}^h B_k F_k(\Gamma^y(z_j))\Gamma^\Omega(z_j) \tag{4.4}$$
where the right hand side is known by step 3. Note that systems derived from even small-scale models may have singular matrix $A^x(z_j)$, so that multiplying through (4.4) by $A^x(z_j)^{-1}$ to obtain $\hat{\Gamma}^x(z_j)$ is not possible. Instead, we multiply through (4.4) by the pseudo inverse of $A^x(z_j)$.

5. **Recursion.** If the relative distance between the guess and updated policy function values is smaller than a pre-specified criterion ϵ, i.e.,
$$\max_{z_1, \ldots, z_N} \frac{\left\|\Gamma^x(z_j) - \hat{\Gamma}^x(z_j)\right\|}{\|\Gamma^x(z_j)\|} < \epsilon$$
where $\|\cdot\|$ denotes some matrix norm, then stop and treat $\{\hat{\Gamma}^x(z_j)\}_{j=1}^N$ as the true policy function values. Otherwise, set $\Gamma^x(x_j) = \hat{\Gamma}^x(z_j)$, $j = 1, 2, \ldots, N$, and go back to step 2.

A distinct advantage of Algorithm 4.2, as we discussed in Section 2.3, is that the iteration centers around the vector of function values $\{(z_j, \Gamma^x(z_j))\}_{j=1}^N$ rather than any particular VARMA parameterization of the true solution. On the other hand, the method of VARMA fitting only serves as an auxiliary tool to facilitate the computation.[19] This approach is different from Sargent (1991) and Huo and Takayama (2018) in that we sidestep the procedure of characterizing the updated VARMA parameterization, which greatly simplifies the computation.

During each iteration of Algorithm 4.2, we solve the model as if agents take the signal structure as an exogenous VARMA process. The resulting data points $\{(z_j, \Gamma^x(z_j))\}_{j=1}^{N}$ then correspond to rational functions even though the true equilibrium solution under endogenous information may not be rational. In this case, theoretically there exist finite orders (p_x, q_x) such that $C^x(z)^{-1}D^x(z) = \Gamma^x(z)$ for all $z \in \mathbb{D}$. Practically, we fix the orders (p_x, q_x) across all iterations of Algorithm 4.2. We find that $(p_x, q_x) = (10, 10)$ are sufficient for most economic applications, though we recommend setting $(p_x, q_x) = (5, 5)$ for small systems. Online Appendix S2 provides computation details of the VARMA fitting method. In our MATLAB toolbox, we also allow for automatic order reduction whenever the fitted VARMA process is non-stationary. This optional feature can be useful for models with non-stationary equilibria, such as the Graham and Wright (2010) model discussed in Section 5.3. We discuss this option in Online Appendix S3.

4.2 Equilibrium Existence and Multiplicity

In Proposition 3 of the appendix, we establish the equilibrium existence and uniqueness for the simplified model system (4.1) under exogenous information. The driving force behind the uniqueness result is that when information is exogenous (including the full-information case), the frequency-domain expectations operators become linear. In this case, the variances of higher-order average expectations are diminishing and bounded by those of lower-order average expectations. When information is endogenous, on the other hand, these operators are in general highly nonlinear. Therefore, one can no longer apply the volatility bounds conditioning on two distinct endogenous information sets. In this case, it is difficult to provide a general characterization for the equilibrium existence and uniqueness.[20]

Endogenous information may create multiple equilibria. When learning from endogenous signals is associated with non-fundamental shocks, self-fulfilling sentiment equilibria can arise as in Acharya, Benhabib and Huo (2021). Multiplicity may also emerge when endogenous signals contain non-invertible (confounding) moving-average components as in Rondina and Walker

[19] As shown in Online Appendix S2, we cast the signal process in a VARMA parametric form to derive its Wold representation using the state-space method.

[20] From a technical point of view, constructing a closed, bounded, and convex self-mapping for $\mathcal{A}$ on the subset of the unbounded space $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ is challenging. Thus, powerful tools like the Schauder-Tychonoff fixed-point theorem cannot be applied. On the other hand, bounded subsets in $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ are normal families in the Montel space of analytic functions, as implied by the Montel theorem. Therefore, one can show that under the given assumptions, $\mathcal{A}$ defines a compact operator with respect to the topology of uniform convergence on compact subsets of $\mathbb{D}$. However, compactness does not hold in general with respect to the norm topology.

(2018).[21] In section 5.3 we provide an example in which multiple equilibria appear as a result of the interaction between the model's primitive structure and its confounding dynamics. Due to the iterative nature, our baseline APFI algorithm does not select among multiple equilibria and may converge to different equilibria depending on the initial conjecture. There is also no guarantee that the baseline APFI algorithm can find an equilibrium if it exists. The selection of equilibria is possible when prior knowledge of the equilibrium structure is available. In a related paper, Adams (2021) studies a class of dispersed-information models under information structure I.1. He characterizes the uniqueness of a "stable" equilibrium class that satisfies the local contraction property within a certain neighborhood. The uniqueness condition in his paper is characterized by the operator norm of the Fréchet derivative of the signal polynomial in the space of Laurent (Toeplitz) operators.

4.3 Extensions

While we present the baseline algorithm based on the simplified system (4.1), our MATLAB toolbox implements the algorithm using the general system (3.1) and such an extension is straightforward. When x_t contains both individual choices x_{1t} and aggregate outcomes x_{2t}, and ϵ_t contains both individual innovations ϵ_{1t} and aggregate innovations ϵ_{2t}, we partition the VMA(∞) representation (3.11) conformably as

$$\underbrace{\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix}}_{x_t} = \underbrace{\begin{bmatrix} \Gamma^x_{11}(\mathrm{L}) & \Gamma^x_{12}(\mathrm{L}) \\ \mathbf{0} & \Gamma^x_{22}(\mathrm{L}) \end{bmatrix}}_{\Gamma^x(\mathrm{L})} \underbrace{\begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}}_{\epsilon_t}.$$

We then solve for the updated policy functions $\Gamma^x_{11}(z)$ and $[\Gamma^x_{12}(z)', \Gamma^x_{22}(z)']'$ in two separate blocks in step 4 of Algorithm 4.2.

In the solving routine of our toolbox, `solve.m`, we offer a comprehensive list of options that help tailor the implementation of Algorithm 4.2 to a specific problem at hand. These options include the initial guess of the policy functions, the convergence criterion, the minimal and maximal numbers of iterations, the VARMA fitting orders, the number of grid points in the discrete Fourier transform, and the updating step size. We find these options valuable in terms of improving the stability and efficiency of our method. The user only needs to express the model in our canonical form and specify the model's information and variable structures—an effort no more complicated than using the DYNARE software. We discuss these implementation details in Online Appendix S3. As will be shown in the next section, the toolbox also allows the user to go

[21] Consider an endogenous signal of the form $p_t = P(\mathrm{L})(\mathrm{L} - \lambda)$, where the confounding dynamics is indexed by $|\lambda| < 1$ and $P(z) \in \mathbf{H}^2(\mathbb{D})$. Multiple equilibria exist when (i) more than one value of λ satisfies the equilibrium restrictions or (ii) there are multiple non-invertible roots in the solution representation of p_t.

beyond the canonical form and call each routine independently when, for example, Assumption 4.1 does not hold.[22]

5 Applications

We provide three macroeconomic models to demonstrate the use of our APFI algorithm and its accuracy, applicability, and flexibility. The first example is a New Keynesian DSGE model with a predetermined physical state and features learning from policies as in Melosi (2017). The second example is the HANK model of Angeletos and Huo (2021). In terms of using the canonical form of our baseline APFI framework, the two examples are similar in that they both feature predetermined variables, and we consider asymmetric and endogenous information sets among agents. In addition, the HANK example allows for endogenous wealth distribution. As a step-by-step user guide, Online Appendix S5 and S6 demonstrate how to cast these models into the canonical form used in our baseline APFI algorithm.

In the final example, we consider the dispersed-information RBC model of Graham and Wright (2010) that violates the regularity condition in Assumption 4.1 and features non-stationary and multiple equilibria. We developed three extended APFI algorithms to solve this model. We also compare their computation performance with the time-domain truncation approach and the closed-form solution. Online Appendix S7 provides detailed documentation on these algorithms along with a user guide for customized coding using our toolbox's individual routines. Unless stated otherwise, all lower case variables in this section have been log-linearized around their steady states.

5.1 DSGE Model with Learning from Policy

5.1.1 Learning from Monetary Policy

The model features a representative household, a continuum of monopolistic competitive intermediate goods firms with Calvo (1983) pricing, a final goods firm, and a monetary authority. Following Melosi (2017), we first assume the household is equipped with full information while the intermediate firms are subject to dispersed information. Melosi (2017) introduces measurement

[22] When Assumption 4.1 is violated, the baseline APFI algorithm can still be applied by recasting the fixed point condition (4.1) as

$$A_0\Gamma^x(z) = -\sum_{k=1}^{l} A_k^x z^k \Gamma^x(z) - \sum_{k=0}^{l} A_k^s z^k \Gamma^s(z) - \sum_{k=0}^{h} B_k F_k(\Gamma^y(z))\Gamma^\Omega(z), \qquad z \in \mathbb{D}$$

Iteration based on the above condition is well-defined and may converge to a stationary solution if the expectational block anchors the explosive behavior in the non-expectational block. However, in practice, the baseline APFI algorithm can be ill-behaved and unstable under this circumstance.

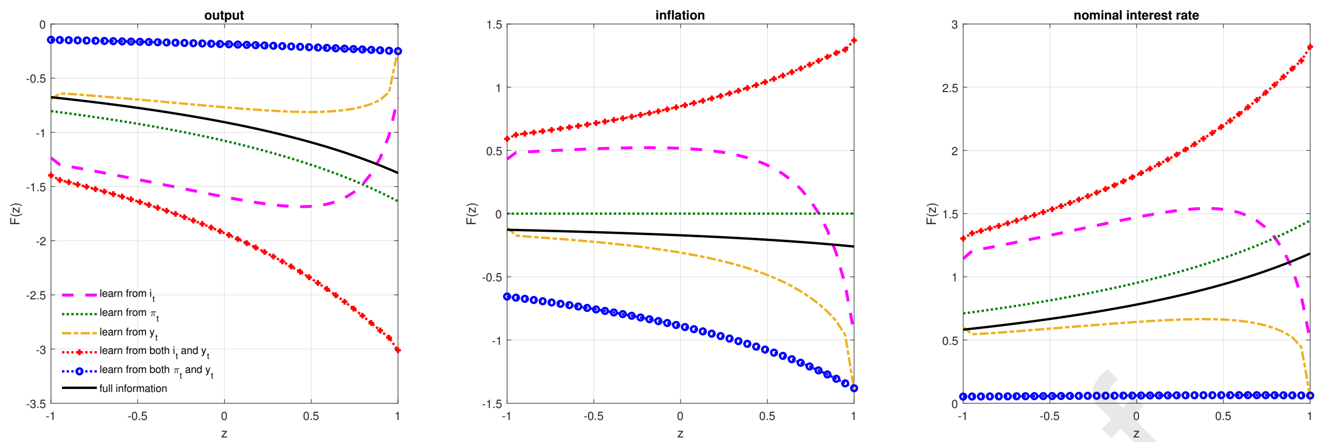

Figure 3: Analytic policy functions to monetary authority's inflation measurement error shock. Parameter values follow Melosi (2017).

error shocks for both inflation π_t and output y_t in the monetary policy rule and highlights the signaling effects of monetary policy. A crucial assumption in his analysis is that the intermediate firms only learn from a single endogenous variable—the nominal interest rate i_t.

This example serves two purposes. First, we illustrate converting a model solution from its analytic function in the frequency domain to its impulse response function in the time domain. Second, our results yield further insights on the sensitivity of the model solution to firms' information sets. Here we consider several different information settings, allowing firms to either learn from i_t, π_t, and y_t each only, from both i_t and π_t, and from both π_t and y_t. We also consider the full information case as a benchmark.[23] Alternative information sets can be easily specified in our MATLAB toolbox by changing a few lines of code. Online Appendix S5 contains the formal model setup.

Figure 3 plots the analytic policy functions of output, inflation, and nominal interest rate to the inflation measurement error shock in the frequency domain under alternative information sets. Different information sets yield strikingly different policy functions (i.e., model solutions). For example, when firms are only learning from i_t as in Melosi (2017), the analytic policy function of inflation has a zero (root) inside the open interval $(-1, 1)$. In contrast, when firms learn from other endogenous variables, the analytic policy functions are either identically zero or have no zeros in $(-1, 1)$.

Figure 4 converts the above analytic policy functions to the impulse response functions in the time domain. When firms are only learning from i_t, the impulse responses of output, inflation, and nominal interest rate to a positive inflation measurement error shock are all hump-shaped, and the inflation response features a price puzzle—a monetary tightening raises the initial price level. By stark contrast, both the prize puzzle and the hump disappear when firms are equipped with

[23]When intermediate firms are learning from all three endogenous variables (i.e., i_t, π_t, and y_t), the model solution becomes indistinguishable from the full-information solution.

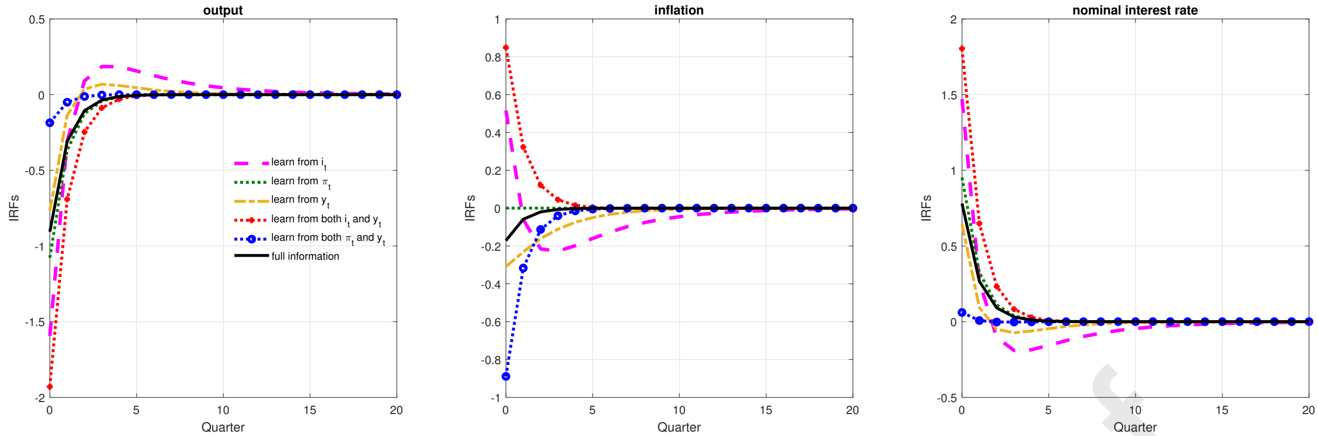

Figure 4: Impulse responses to a positive, one standard deviation inflation measurement error shock. Parameter values follow Melosi (2017).

full information. The hump-shaped impulse responses to a contractionary monetary shock is a universal finding that is robust to many identification schemes [see, e.g., Christiano, Eichenbaum and Evans (1999)]. While a thorough investigation of the impacts of different information sets is beyond the scope of this paper, our results suggest that information frictions may be as important as many other commonly used real and nominal rigidities in determining a DSGE model's empirical performance.

5.1.2 Learning from Fiscal Policy

The recent COVID-19 pandemic has been a substantial shock to the U.S. economy. In response to the significant public health and economic crisis, the federal government has conducted a series of fiscal expansions. The unprecedented large-scale fiscal stimulus, however, raises concerns about its impact on inflation. In a recent article, Larry Summers (2021) warned that the fiscal stimulus "will set off inflation pressures of a kind we have not seen in a generation". One day later, Paul Krugman (2021) took a different approach and argued that "even a very hot economy only leads to modest inflationary overheating". We offer novel insight into this ongoing debate by considering how fiscal policy affects inflation under various information structures. We also use this example to demonstrate how the APFI framework handles multiple types of incomplete information sets.

We first augment the dispersed-information DSGE model of Section 5.1.1 by a simple fiscal sector. The fiscal authority imposes lump-sum net taxes T_t and issues one-period nominal bond B_t. Let B_t/P_t denote the real debt where P_t is the price level. The fiscal authority's primary surplus is defined by $S_t = T_t - G_t$. We assume there is no government spending, i.e., $G_t \equiv 0$.

The government's flow budget constraint is given by

$$\frac{1}{R_t}\frac{B_t}{P_t} + S_t = \frac{B_{t-1}}{P_t} \tag{5.1}$$

To close the fiscal sector, we adopt a simple rule for the primary surplus. Let s_t, b_t denote log-deviations of $S_t, B_t/P_t$ from their steady state values. The fiscal rule is of the form

$$s_t = \gamma b_{t-1} + \xi_{s,t}, \qquad \xi_{s,t} = \rho_s \xi_{s,t-1} + \eta_t^s, \qquad \eta_t^s \sim \mathbb{N}(0, \sigma_s^2) \tag{5.2}$$

where the parameter γ governs how aggressively the primary surplus responds to the lagged debt.

Let the parameter ϕ_π govern the responsiveness of nominal interest rate R_t to inflation π_t. Our subsequent analysis focuses on the active monetary ($\phi_\pi > 1$) and passive fiscal ($\gamma > 1$) policy regime [see Leeper (1991)], which implies a Ricardian fiscal policy under full information. In particular, given lump-sum taxation, the impulse responses of output to fiscal innovations (i.e., η_t^s) are trivially zero. Moreover, once we shut down all non-policy shocks, inflation is a purely monetary phenomenon. As the impulse responses of inflation to η_t^s are also zero, changes in the size of government debt b_t have no impact on inflation.

The dispersed-information DSGE model of Section 5.1.1 permits incomplete-information firms. We enrich the analysis here by allowing for an incomplete-information representative household as well. Furthermore, we allow households and firms to observe differential, non-nested, and endogenous information. To the best of our knowledge, the extension of introducing incomplete information to both the supply and demand sides is a novel contribution to the literature. It also highlights the flexibility of our methodology and numerical toolbox.

We maintain the key message of Melosi (2017) that policy variables can serve as endogenous signals to the private sector. There are three policy variables—the nominal interest rate i_t, the primary surplus s_t, and the real debt b_t. To focus on the implications of the fiscal signals, we include the history of i_t in both the household's and firms' information sets. We consider four cases where only one fiscal variable (either s_t or b_t) enters either the representative household's or the intermediate firms' incomplete information sets.[24] That is,

$$\begin{aligned}
&\text{Case 1:} \quad \mathcal{I}_t^{HH} = \{i_{t-j} : j \geqslant 0\}, \quad \mathcal{I}_{t,i}^{Firm} = \{i_{t-j}, s_{t-j}, a_{t-j}^i : j \geqslant 0\} \\
&\text{Case 2:} \quad \mathcal{I}_t^{HH} = \{i_{t-j} : j \geqslant 0\}, \quad \mathcal{I}_{t,i}^{Firm} = \{i_{t-j}, b_{t-j}, a_{t-j}^i : j \geqslant 0\} \\
&\text{Case 3:} \quad \mathcal{I}_t^{HH} = \{i_{t-j}, s_{t-j} : j \geqslant 0\}, \quad \mathcal{I}_{t,i}^{Firm} = \{i_{t-j}, a_{t-j}^i : j \geqslant 0\} \\
&\text{Case 4:} \quad \mathcal{I}_t^{HH} = \{i_{t-j}, b_{t-j} : j \geqslant 0\}, \quad \mathcal{I}_{t,i}^{Firm} = \{i_{t-j}, a_{t-j}^i : j \geqslant 0\}
\end{aligned}$$

[24] If neither primary surplus s_t nor real debt b_t enters the households' and firms' information sets, then the Ricardian equivalence of fiscal policy still holds. On the other hand, if primary surplus enters one private sector's information set while real debt enters the other private sector's information set, the fiscal impulse responses are quantitatively small, indicating small deviations from the Ricardian fiscal policy.

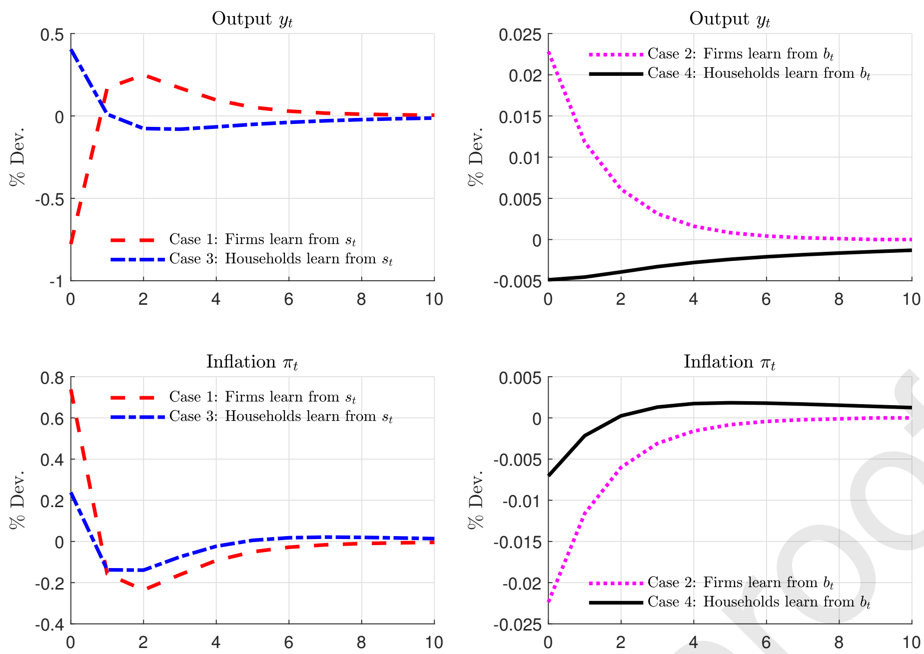

Figure 5: Impulse responses (% deviations) of output and inflation to one standard deviation decrease in the primary surplus shock (i.e., $\eta_0^s = -\sigma_s$) under different information environments. Non-fiscal parameter values follow Melosi (2017). Fiscal parameter values follow the high-pass posterior estimates of Tan (2019).

These cases can be easily specified in our toolbox by letting various variables enter different agents' information sets.

It is worth noting that when a fiscal variable enters the household's information set (i.e., Case 3 and Case 4), the two private sectors' information sets become non-nested. That is, $\mathcal{I}_t^{HH} \not\subset \mathcal{I}_{t,i}^{Firm}$ and $\mathcal{I}_{t,i}^{Firm} \not\subset \mathcal{I}_t^{HH}$. Non-nested information sets further complicate the issue of higher-order expectations (HOEs). Since

$$\text{Case 1 and 2:} \quad \mathbb{E}_t^{HH}\mathbb{E}_{t,i}^{Firm}\pi_t = \mathbb{E}_{t,i}^{Firm}\mathbb{E}_t^{HH}\pi_t = \mathbb{E}_t^{HH}\pi_t$$
$$\text{Case 3 and 4:} \quad \mathbb{E}_t^{HH}\mathbb{E}_{t,i}^{Firm}\pi_t \neq \mathbb{E}_{t,i}^{Firm}\mathbb{E}_t^{HH}\pi_t \neq \mathbb{E}_t^{HH}\pi_t,$$

non-nested information sets introduce additional HOEs. Time-domain methodologies require the inclusion of a large number of HOEs to form a suitable state space. A direct consequence of non-nested information sets is that time-domain methods will suffer from the curse of dimensionality even more. More importantly, the underlying law of motion for HOEs may not take the simple VAR(1) form as typically postulated in the existing time-domain methods [see, e.g., Nimark (2008)].

Figure 5 plots the impulse responses of output y_t and inflation π_t to one standard deviation decrease in the primary surplus shock (i.e., $\eta_0^s = -\sigma_s$). Different information sets generate qualitatively different initial responses of output and inflation at $t = 0$. Compared to the full-information model, both output and inflation display non-trivial responses in all cases. The fiscal shock could either expand or contract the real output and be either inflationary or deflationary. Why are the impacts of fiscal shocks qualitatively different under incomplete information? Besides the conventional roles of policy instruments, these policy variables also serve as endogenous signals that reveal the economic fundamentals. Different perceived fundamentals affect the private sectors' expectation formation and decision making, thereby generating strikingly different impulse responses.

Comparing the left and right panels of Figure 5 suggests that the fiscal impacts also differ quantitatively. The most undesirable case may be Case 1, where output drops the most on impact, and inflation increases significantly. These effects quickly reverse themselves in the subsequent periods. Interestingly, under the current parameterization, both output and inflation deviate little from zero as long as one private sector (either the household or firms) learns from the real debt b_t. Since the real value of government debt must be equal to the present value of current and expected future primary surpluses, knowing the history of real debt provides the private sector with much information about the history of primary surpluses. As one private sector nearly figures out the fiscal shock, the resulting fiscal impacts deviate little from its full-information, Ricardian benchmark through the general equilibrium effect, even though the other private sector pays no attention to b_t.

5.2 HANK Model with Incomplete Information

Next we consider the incomplete-information heterogeneous agent New Keynesian (HANK) model of Angeletos and Huo (2021) with endogenous wealth distribution. This model features consumer heterogeneities in business cycle exposure and marginal propensity to consume (MPC). There are two groups of consumers, indexed by $g = \{1, 2\}$ with respective mass $\pi_g = 0.5$. Let w_g denote the survival rate of individuals of each group in each period, and ϕ_g denote the exposure to business cycles. The (log) income of group g is $y_{g,t} = \phi_g y_t$, where $\phi_g \geqslant 0$ is the elasticity of group g's income with respect to the aggregate income (i.e., output) and $\pi_1\phi_1 + \pi_2\phi_2 = 1$. The MPC of each group is given by $1 - \beta w_g$. In the model, $w_1 < w_2$ and $\phi_1 > \phi_2$. Consequently, group 1 consumers are subject to both high cyclical exposure and high MPC. The formal model setup is contained in Online Appendix S6.

Denote the group-level consumption and saving as $c_{g,t}$ and $s_{g,t}$, respectively.[25] As shown in Online Appendix S6, the group level consumption can be expressed as

$$c_{g,t} = (1-\beta w_g)\frac{1}{\beta}s_{g,t-1} - \beta w_g \sum_{j=0}^{\infty}(\beta w_g)^j \overline{\mathbb{E}}_{g,t}[r_{t+j}] + (1-\beta w_g)\phi_g \sum_{j=0}^{\infty}(\beta w_g)^j \overline{\mathbb{E}}_{g,t}[y_{t+j}] \tag{5.3}$$

[25] The lower case variable $s_{g,t}$ stands for the ratio between the group saving level $S_{g,t}$ and the natural level of output Y^* (i.e., $s_{g,t} = \frac{S_{g,t}}{Y^*}$) as the steady state value of $S_{g,t}$ is zero.

where $\overline{\mathbb{E}}_{g,t}[\cdot]$ is the average expectation of group g at time t, and r_t is the unobserved (log) real interest rate subject to exogenous shocks. Each individual i in either group g forms expectations $\mathbb{E}_{i,g;t}[\cdot]$ conditional on a private, noisy signal on the real interest rate. The group level budget constraint is

$$c_{g,t} + s_{g,t} = \frac{1}{\beta} s_{g,t-1} + \phi_g y_t \tag{5.4}$$

and the market clearing condition requires

$$\pi_1 s_{1,t} + \pi_2 s_{2,t} = 0 \tag{5.5}$$

The saving pair $(s_{1,t}, s_{2,t})$ defines the endogenous wealth distribution among groups in the economy. In the majority of their analysis, Angeletos and Huo (2021) impose a sequence of fiscal transfers to undo any wealth inequality triggered by the interest rate shocks so that $s_{g,t} \equiv 0$. They then provide intuitions on why allowing endogenous wealth dynamics adds persistence to the aggregate output in response to the interest rate shocks.

The HANK model serves three purposes. First, as a practical guide, we illustrate how to introduce dummy variables to map the equilibrium condition (5.3), which involves an infinite sum of expectations, into the canonical form. Infinite sums of expectations frequently appear in incomplete-information models with an infinite number of agents and higher-order expectations. Second, we offer additional insight into the extra persistence due to endogenous wealth distribution. Third, we illustrate the robustness of the results by introducing an asymmetric, endogenous signal.

To begin with, define

$$x_{i,g;t} = \beta w_g \mathbb{E}_{i,g;t} \sum_{j=0}^{\infty} (\beta w_g)^j r_{t+j}, \qquad z_{i,g;t} = (1-\beta w_g)\phi_g \mathbb{E}_{i,g;t} \sum_{j=0}^{\infty} (\beta w_g)^j y_{t+j}$$

Since the law of iterated expectations applies to individual expectations $\overline{\mathbb{E}}_{g,t}[\cdot]$, Online Appendix S6 shows we can rewrite the dummy variables $x_{i,g;t}, z_{i,g;t}$ recursively as

$$x_{i,g;t} = \beta w_g \mathbb{E}_{i,g;t}\left[r_t\right] + \beta w_g \mathbb{E}_{i,g;t+1}\left[x_{i,g;t+1}\right] \tag{5.6}$$

$$z_{i,g;t} = (1-\beta w_g)\phi_g \mathbb{E}_{i,g;t}\left[y_t\right] + \beta w_g \mathbb{E}_{i,g;t+1}\left[z_{i,g;t+1}\right] \tag{5.7}$$

Using the two dummy variables, the equilibrium condition (5.3) can be expressed as

$$c_{g,t} = (1-\beta w_g)\frac{1}{\beta} s_{g,t-1} - \int_{[0,1]} x_{i,g;t} di + \int_{[0,1]} z_{i,g;t} di \tag{5.8}$$

One can then map (5.4)–(5.8) into the canonical form and solve the model using our toolbox.

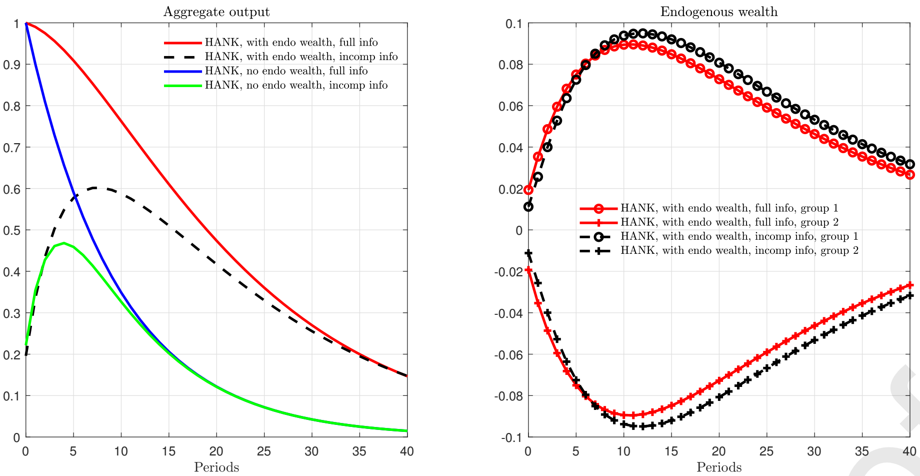

Figure 6: Impulse responses of aggregate output (left panel) and endogenous wealth (right panel) to a negative real interest rate shock. The shock magnitude is normalized so that the full information response of aggregate output on impact equals to 1. Parameter values follow Angeletos and Huo (2021).

Figure 6 plots the impulse responses of the aggregate output and the endogenous wealth distribution to a negative real interest rate shock under both full and incomplete information. Interestingly, while there is a significant distinction between the impulse responses of the aggregate output (i.e., monotone vs. hump-shaped), the wealth distributions display similar patterns between the full-information and the incomplete-information models. We also consider the case of eliminating the endogenous wealth inequality. Consistent with Angeletos and Huo (2021), the impulse responses of the aggregate output are much more persistent when the endogenous wealth inequality is allowed under both full and incomplete information.

Online Appendix S6 shows that when the wealth inequality is allowed, the aggregate output follows

$$y_t = \frac{1}{1-\theta\,\mathrm{L}}\left[\pi_1(1-w_1\,\mathrm{L})\int_{[0,1]}(z_{i,1;t}-x_{i,1;t})\,di + \pi_2(1-w_2\,\mathrm{L})\int_{[0,1]}(z_{i,2;t}-x_{i,2;t})\,di\right] \tag{5.9}$$

where $\theta = \pi_1\phi_1 w_1 + \pi_2\phi_2 w_2 < 1$. In contrast, when there is no wealth inequality, the aggregate output is given by

$$y_t = \pi_1\int_{[0,1]}(z_{i,1;t}-x_{i,1;t})\,di + \pi_2\int_{[0,1]}(z_{i,2;t}-x_{i,2;t})\,di \tag{5.10}$$

Without endogenous wealth, (5.10) defines the aggregate demand of the economy as a dynamic network among the two groups of consumers. Comparing (5.9) with (5.10) indicates additional autoregressive and moving-average terms arise when the wealth inequality is allowed, both of which contribute to the higher persistence of the aggregate output.

We now enrich the HANK model by introducing a noisy private signal $m_{i,g;t}$ on the endogenous

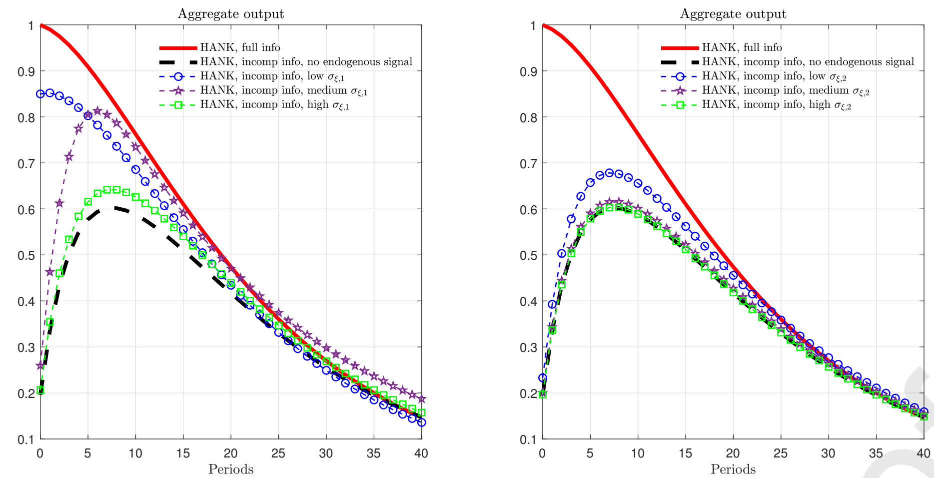

Figure 7: Impulse responses of aggregate output to a negative real interest rate shock. The shock magnitude is normalized so that the full information response on impact equals to 1. Parameter values other than $\sigma_{\xi,g}$ follow Angeletos and Huo (2021).

aggregate output y_t.[26] The group-specific signal $m_{i,g;t}$ is of the form

$$m_{i,g;t} = y_t + \xi_{i,g;t}, \qquad \xi_{i,g;t} \sim \mathbb{N}(0, \sigma^2_{\xi,g}) \tag{5.11}$$

We assume only one group receives such a signal. While Angeletos and Huo (2021) mainly focus on symmetric and exogenous information, we allow asymmetric and endogenous information, and focus on how such an information setting impacts the aggregate output dynamics. Figure 7 plots the impulse responses of y_t when either group 1 (left panel) or group 2 (right panel) receives the endogenous signal $m_{i,g;t}$. We vary $\sigma_{\xi,g}$ and consider an $m_{i,g;t}$ with a high, medium, and low precision.[27] Figure 7 also plots the two benchmark cases where both groups are equipped with either full information or symmetric, exogenous information. All models considered in Figure 7 allow for endogenous wealth distribution.

Comparing the left and right panels of Figure 7 indicates that information asymmetry matters in shaping the aggregate dynamics. Providing group 1 with a precise endogenous signal (i.e., low $\sigma_{\xi,1}$) yields a monotone impulse response that is quite similar to its full-information counterpart.[28] In contrast, providing group 2 with the same signal (i.e., low $\sigma_{\xi,2}$) generates a hump-shaped impulse response. When the magnitude of $\sigma_{\xi,2}$ is medium or high, the impulse responses are

[26] In the toolbox, we also include the customized solution code for the HANK model that does not rely on the canonical form. In that example, we instead consider a noisy, group-specific aggregate signal. The discounted infinite future sum of expectations can be handled directly by a separate routine. The results are similar.

[27] Let $\sigma^f(y_t)$ denote the unconditional standard deviation of the aggregate output in the full-information HANK model with endogenous saving. We pick low, medium, and high values of $\sigma_{\xi,g}$ such that $\sigma_{\xi,g}/\sigma^f(y_t) \in \{0.01, 0.25, 1\}$.

[28] Increasing $\sigma_{\xi,1}$ from low to medium value generates a pronounced hump-shaped response. Interestingly, the impulse response under-shoots initially but then over-shoots its counterpart under full information. Angeletos, Huo and Sastry (2020) emphasize a similar pattern found in surveys of macroeconomic expectations.

indistinguishable from its incomplete-information counterpart with symmetric and exogenous signals.

These results suggest the endogenous signal $m_{i,2;t}$ about the aggregate output y_t does not change the behavior of group 2 individuals much when it is not precise enough. In the extreme case of $\phi_2 = 0$ so that the income of group 2 individuals is not subject to business cycle fluctuations, the last term in (5.3) vanishes and the group level consumption $c_{2,t}$ does not depend on y_t explicitly. The signal $m_{i,2;t}$ is still useful to group 2 individuals as it contains information about the interest rate shock through y_t. Nevertheless, such a signaling channel only manifests when $\sigma_{\xi,2}$ is small enough. When $\sigma_{\xi,2}$ is relatively large, the pattern of the impulse responses in Figure 7 supports rational inattention of group 2 individuals.

On the other hand, group 1 individuals are subject to both high MPC and high cyclical exposure. The combined features imply a volatile group level consumption $c_{1,t}$. Such a positive correlation between MPC and cyclical exposure also suggests group 1 individuals should pay close attention to business cycle conditions if allowed. If we interpret the negative interest rate shock as a result of an expansionary monetary policy and the existence of $m_{i,1;t}$ due to imperfect central bank communication, Figure 7 suggests the monetary policy effects depend crucially on (i) whether group 1 individuals are learning from $m_{i,1;t}$, and (ii) whether the communication is effective (i.e., a small $\sigma_{\xi,1}$).

5.3 Beyond the Canonical Form and Baseline Algorithm

In the last example, we study an island-type stochastic growth model with dispersed information in Graham and Wright (2010). This model features non-stationarity and multiple equilibria. We go beyond the canonical form and demonstrate how the user can tailor a model-specific APFI algorithm. There are a large number of islands in the economy. Identical households and firms live on each island i, and are subject to unobservable aggregate and idiosyncratic productivity shocks denoted as a_t and z_{it}, respectively. Both shocks follow AR(1) processes. Households consume a single good c_{it}, supply labor n_{it} to local firms, and save in terms of the physical capital k_{it}. Households lease capital at the rental rate r_{kt} to firms on different islands in a centralized capital market, while labor markets are segmented on each island (i.e., labor is immobile). Households and firms on each island share dispersed information about the unobserved state of the economy (i.e., aggregate and idiosyncratic productivity shocks and the aggregate capital). The information set on each island contains two signals: (i) an island-specific wage signal that is informationally equivalent to an exogenous signal as the sum of the aggregate and idiosyncratic productivity shocks, i.e., $s_{wt} = a_t + z_{it}$; (ii) the endogenous rental rate of capital r_{kt}.

Under the parameter calibration in Graham and Wright (2010), the model's equilibrium system does not satisfy the invertibility condition in Assumption 4.1. On the other hand, the number of signals in the model equals the number of unobserved exogenous shocks, and the (endogenous) capital rental rate signal displays confounding dynamics. As such, we solve and characterize the closed-form equilibrium using the frequency-domain techniques similar to Rondina and Walker (2018). The closed-form characterization shows that the full-information and the incomplete-information models cannot be fully stationary. Two possible equilibria are stationary at the aggregate level. The first equilibrium (Equilibrium 1) features a unit root in the idiosyncratic consumption and capital. The second equilibrium (Equilibrium 2) features an explosive process for the idiosyncratic capital. Graham and Wright (2010) focused on the exploration of Equilibrium 1. We refer readers to Online Appendix S7 for the formal model setup and complete characterization of the model solution.

The underlying irregularities impose substantial challenges to numerical computation. In this regard, we reduce the model's equilibrium system into a two-equation system.

$$\begin{cases} c_{it} = \mathbb{E}_{it}\left[c_{it+1} - (1-\beta(1-\delta))r_{kt+1}\right] \\ Q(\mathrm{L})r_{kt} = P_a(\mathrm{L})a_t - P_c(\mathrm{L})c_t \end{cases} \tag{5.12}$$

which consists of the individual Euler equation and a relation between the capital rental rate and the aggregate consumption. The individual expectation is conditional on the information set $X_{it} = \{r_{k,t-j}, s_{w,t-j} : j \geqslant 0\}$. The $P_a(\mathrm{L})$, $P_c(\mathrm{L})$, and $Q(\mathrm{L})$ defined in (S7.14)–(S7.16) of Online Appendix S7 are first-order exogenous lag polynomials associated with the model's primitive parameters. Non-invertibility appears in the second equation of (5.12) as the roots of $Q(z)$ and $P_c(z)$ are all inside the unit circle.[29] Without further restrictions, a well-defined stationary process for c_t could lead to non-stationarity in r_{kt} and vice versa. This problem is independent of the model's information structure and creates numerical instability in an iterative algorithm. In Online Appendix S7 we discuss the origin and implications of the non-invertibility problem in more details.

Motivated by this model's theoretical and numerical irregularities, we develop three variants of the baseline APFI algorithm beyond the canonical form. These algorithms are designed to solve the system (5.12), which is sufficient to characterize the entire model solution. The first algorithm (Algorithm 1) employs the signal representation of the solution to eliminate the non-invertibility problem. For Equilibrium 1 with a random walk, we adopt a first-differencing strategy. The second algorithm (Algorithm 2) expands conditional expectations using the Wiener-Hopf prediction formula and iterates on the functions representing the endogenous signal along with its Wold representation. The third algorithm (Algorithm 3) imposes a functional form restriction for the solution and updates the iteration based on the Euler equation errors, ensuring stationarity during the iteration process. In some of these extended APFI algorithms, we utilize knowledge

[29] If one of the roots of $Q(z)$ and $P_c(z)$ are outside the unit circle, then rearranging the order of variables could eliminate the non-invertibility problem.

Performance	Extended APFI			Time-Domain
Statistics	Algorithm 1	Algorithm 2	Algorithm 3	Truncation
PI Equilibrium 1				
Computation time	6.3691(s)	1.2587(s)	N/A	0.3818(s)
Convergence criterion	10^{-6}	10^{-6}	N/A	10^{-5}
Initial conjecture	Generic	Generic	N/A	Restrictive
PI Equilibrium 2				
Computation time	21.6737(s)	0.8898(s)	531.4262(s)	N/A
Convergence criterion	10^{-6}	10^{-6}	10^{-6}	N/A
Initial conjecture	Generic	Generic	Generic	N/A

Table 2: Statistics of computation performance. N/A denotes the case where the algorithm is not applicable. "Generic" initial conjecture means the initial function values can be set to simple guess, including 0, $\frac{1}{(1-\rho_a)z}$, $-z$, where ρ_a is the shock persistence. On the other hand, "restrictive" initial conjecture means the convergence and stability of the algorithm are highly sensitive to the initial guess.

about the theoretical properties of the model's equilibrium to impose restrictions on the solution form and the functional constants associated with agents' conditional expectations. Such prior knowledge includes the stationarity property of the two equilibria and the structure of the full-information solution.[30]

We conducted a comparison exercise using our APFI algorithms and the time-domain truncation algorithm used in Graham and Wright (2010).[31] In contrast to the time-domain method, we find that the extended APFI algorithms are, in general, slower but more accurate and robust to initial conjectures. Unlike the time-domain approach, our APFI algorithms can also compute different types of equilibria that this model admits. We summarize the computation performance of these algorithms in Table 2. The performance statistics include computing time (in seconds), convergence criterion, and initial conjecture. We report the statistics for the computation of two incomplete-information (PI) equilibria: Equilibrium 1 and Equilibrium 2. We also report the statistics for the time-domain truncation method of Nimark (2017) that is used in Graham and Wright (2010).[32]

[30]If there is no unit root in the system, Algorithm 1 based on the signal representation requires no additional restrictions. See Algorithm S7.8 in Online Appendix S7.

[31]We are grateful to Liam Graham and Stephen Wright for sharing their original code with us for completing this exercise.

[32]The numerical experiment is conducted on a laptop with Intel® Core™ i5-4600U CPU (2 cores, 4 threads), 2.40GHz, RAM 8.00Gb. We run the code on Matlab R2019b platform using individual routines contained in our toolbox. For each statistic, we run the algorithm ten times and calculate the average performance. We use the same parameter setting as in Graham and Wright (2010) except that we set the steady-state growth rate to $g = 0$ in each algorithm. However, all results remain valid when we set $g > 0$. In the toolbox, we also include the code for solving the full-information equilibrium, and its numerical performance is similar to the PI case.

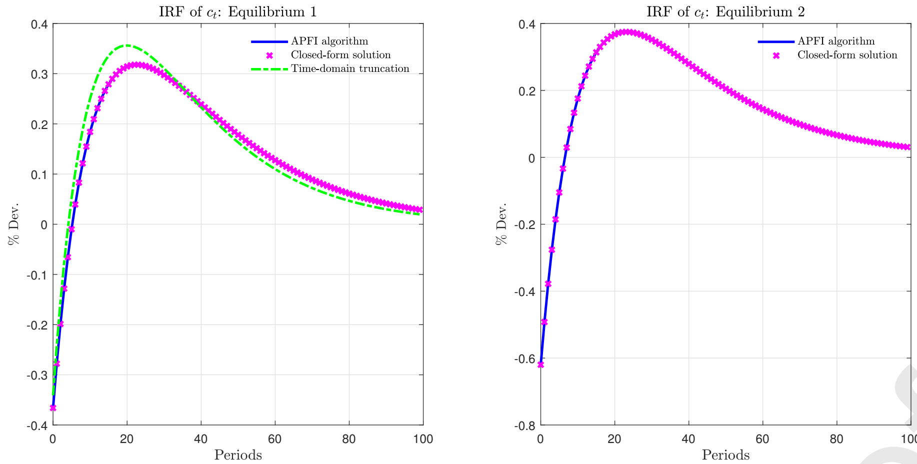

Figure 8: Comparison of frequency-domain and time-domain numerical solutions with closed-form solution under incomplete information.

As Table 2 suggests, the APFI algorithms turn out to be slower than the time-domain truncation method. This is not surprising considering the amount of algebra involved in the latter approach. The APFI algorithms, on the other hand, entail a series of intermediate steps (e.g., finding the rational approximation, computing the Wold spectral factorization, and evaluating the annihilation operator), which are performed by the individual routines in our toolbox. Among all, Algorithm 3 requires the most significant amount of algebraic work and hence is the slowest. However, various options can be applied to improve the speed of our APFI algorithms further, including reducing the grid points on the state space, adopting more delicate initial guesses, reducing the number of points for the discrete Fourier transform, and reducing the orders of VARMA(p, q) approximation. In Online Appendix S7 we provide some helpful coding pointers that guide these implementation details.

Compared to the performance on speed, we believe the accuracy, stability, and flexibility of an algorithm are equally important. In these regards, the APFI algorithms demonstrate superior advantages. The left panel of Figure 8 plots the impulse response of the aggregate consumption in PI Equilibrium 1 under different APFI algorithms, the closed-form solution, and that under the time-domain algorithm. The APFI algorithms lead to a highly accurate solution virtually identical to the true solution. On the other hand, the time-domain algorithm used by Graham and Wright (2010) yields a substantial numerical error, which remains large even if we increase the order of higher-order expectations in the state space. In addition, we identify and compute the PI Equilibrium 2, which displays a qualitatively similar solution. Quantitatively, the two equilibria are sufficiently different. The right panel of Figure 8 plots the impulse response of the aggregate consumption in Equilibrium 2 under the APFI algorithms and the true solution. Again, our algorithms demonstrate superb accuracy. Finally, all three APFI algorithms are robust to different initial conjectures as generic initial guesses lead to stable convergence. This robustness is in sharp contrast to the time-domain approach, whose convergence is sensitive to the initial guess.[33]

The dynamic responses in Figure 8 demonstrate high persistence even when the shock persistence is only moderate (0.9). The impulse response sequences (i.e., MA(∞) coefficients of the solution) do not vanish even after 100 periods. This observation implies that the MA truncation methods such as in Lorenzoni (2009) and more recently in Adams (2021) may become less efficient as the required truncation length and the number of unknown coefficients that need to be solved are enormous.

6 Concluding Remarks

We have developed a unified framework for solving and analyzing dynamic macroeconomic and finance models of incomplete information. In particular, we propose a policy function iteration method based on the frequency-domain techniques and provide a handy numerical toolbox that implements our method. We also demonstrate the applicability and flexibility of this framework using several economic examples. We believe the tools developed in this paper can be useful in addressing many potential research questions. Extending our methodology to continuous-time models or even to nonlinear models will also be important future directions.

References

Acharya, S. (2014): "Dispersed Beliefs and Aggregate Demand Management," Working Paper.

Acharya, S., J. Benhabib, and Z. Huo (2021): "The anatomy of sentiment-driven fluctuations," *Journal of Economic Theory*, 195, 105280.

Adams, J. J. (2021): "Macroeconomic Models with Incomplete Information and Endogenous Signals," Manuscript.

Al-Sadoon, M. M. (2018): "The Linear Systems Approach to Linear Rational Expectations Models," *Econometric Theory*, 34(3), 628–658.

——— (2020): "The Spectral Approach to Linear Rational Expectations Models," Manuscript.

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

[33] In the code of Graham and Wright (2010), they first compute a "pseudo model" and then use the resulting solution as an initial guess. However, their algorithm convergence is vulnerable to other simpler initial guesses.

Angeletos, G.-M., Z. Huo, and K. A. Sastry (2020): *Imperfect Macroeconomic Expectations: Evidence and Theory*pp. 1–86. University of Chicago Press.

Angeletos, G.-M., and C. Lian (2016): "Chapter 14 - Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination," vol. 2 of *Handbook of Macroeconomics*, pp. 1065 – 1240. Elsevier.

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

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

Beliaev, D. (2019): *Conformal Maps and Geometry*. WORLD SCIENTIFIC (EUROPE).

Blanchard, O. J., J.-P. L'Huillier, and G. Lorenzoni (2013): "News, Noise, and Fluctuations: An Empirical Exploration," *American Economic Review*, 103(7), 3045–70.

Calvo, G. A. (1983): "Staggered Prices in a Utility Maxmimizing Model," *Journal of Monetary Economics*, 12(3), 383–398.

Chahrour, R., and G. Gaballo (2019): "Learning from House Prices: Amplification and Business Fluctuations," CEPR Discussion Paper No. DP14120.

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

Chahrour, R., and R. Ulbricht (2018): "Information-Driven Business Cycles: A Primal Approach," Working Paper.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Chapter 2 Monetary policy shocks: What have we learned and to what end?," vol. 1 of *Handbook of Macroeconomics*, pp. 65 – 148. Elsevier.

Conway, J. (1978): *Functions of One Complex Variable I*, Functions of one complex variable / John B. Conway. Springer.

Conway, J. B. (1990): *A Course in Functional Analysis*, Springer Graduate Texts in Mathematics. Springer.

Duren, P. (2000): *Theory of Hp Spaces*, Dover books on mathematics. Dover Publications.

Forni, M., L. Gambetti, and L. Sala (2019): "Structural VARs and noninvertible macroeconomic models," *Journal of Applied Econometrics*, 34(2), 221–246.

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.

Han, Z., X. Ma, and R. Mao (2019): "Roles of Dispersed Information on Inflation and Inflation Expectations," .

Hansen, L. P., and T. J. Sargent (1980): "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2, 7–46.

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

Huo, Z., and N. Takayama (2018): "Rational Expectations Models with Higher Order Beliefs," Working Paper.

Janashia, G., E. Lagvilava, and L. Ephremidze (2011): "A New Method of Matrix Spectral Factorization," *IEEE Transactions on Information Theory*, 57(4), 2318–2326.

Judd, K. L. (1998): *Numerical Methods in Economics*. The MIT Press, Cambridge, MA.

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," *The Review of Economic Studies*, 81(3), 1137–1163.

Katznelson, Y. (1976): *An Introduction to Harmonic Analysis*, Dover books on advanced mathematics. Dover Publications.

Kohlhas, A. N. (2019): "Learning by Sharing: Monetary Policy and Common Knowledge," *American Economic Journal: Macroeconomics*.

Krugman, P. (2021): "Stagflation revisited," https://paulkrugman.substack.com/p/stagflation-revisited.

Leeper, E. M. (1991): "Equilibria Under 'Active' and 'Passive' Monetary and Fiscal Policies," *Journal of Monetary Economics*, 27(1), 129–147.

Lindquist, A., and G. Picci (2015): *Linear Stochastic Systems: A Geometric Approach to Modeling, Estimation and Identification*, Series in Contemporary Mathematics. Springer Berlin Heidelberg.

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

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

MELOSI, L. (2017): "Signalling Effects of Monetary Policy," *The Review of Economic Studies*, 84(2), 853–884.

MIAO, J., J. WU, AND E. YOUNG (2021a): "Multivariate Rational Inattention," Working Paper.

MIAO, J., J. WU, AND E. R. YOUNG (2021b): "Macro-financial volatility under dispersed information," *Theoretical Economics*, 16(1), 275–315.

NIMARK, K. (2008): "Dynamic pricing and imperfect common knowledge," *Journal of Monetary Economics*, 55(2), 365 – 382.

——— (2017): "Dynamic Higher Order Expectations," CEPR Discussion Papers No. DP11863.

OPPENHEIM, A. V., A. S. WILLSKY, AND I. T. YOUNG (1983): *Signals and Systems*, Prentice-Hall Signal-Processing Series. Prentice-Hall.

RONDINA, G., AND T. B. WALKER (2018): "Confounding Dynamics," Working Paper.

ROZANOV, Y. A. (1967): *Stationary Random Processes*. Holden-Day, San Francisco.

RUDIN, W. (1987): *Real and Complex Analysis*, Mathematics series. McGraw-Hill.

SARGENT, T. J. (1991): "Equilibrium with Signal Extraction from Endogenous Variables," *Journal of Economic Dynamics and Control*, 15, 245–273.

SAYED, A. H., AND T. KAILATH (2001): "A survey of spectral factorization methods," *Numerical Linear Algebra with Applications*, 8(6?7), 467–496.

SIMS, C. A. (1992): "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975–1000.

SINGLETON, K. J. (1987): "Asset Prices in a Time Series Model with Disparately Informed, Competitive Traders," in *New Approaches to Monetary Economics*, ed. by W. Barnett, and K. Singleton. Cambridge University Press, Cambridge.

SUMMERS, L. (2021): "The Biden stimulus is admirably ambitious. But it brings some big risks, too," https://www.washingtonpost.com/opinions/2021/02/04/larry-summers-biden-covid-stimulus.

Tan, F. (2019): "A Frequency-domain Approach to Dynamic Macroeconomic Models," *Macroeconomic Dynamics*, pp. 1–31.

Tan, F., and T. B. Walker (2015): "Solving Generalized Multivariate Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 60, 95–111.

Tang, J. (2015): "Uncertainty and the Signaling Channel of Monetary Policy," FRB of Boston-Working Paper.

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

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

Whiteman, C. H. (1985): "Spectral utility, wiener-hopf techniques, and rational expectations," *Journal of Economic Dynamics and Control*, 9(2), 225 – 240.

Appendix

Proof of Theorems

This section presents the proof of three theorems in the main text.

Proof of Theorem 3.1

We start with a subset of $\bigcup_{p,q\in\mathbb{N}} \mathbf{Q}_{(p,q)}$ that corresponds to VARMA$(k, k-1)$ processes. In particular, define the set of $n_x \times n_\epsilon$ matrices of proper rational analytic functions that correspond to VARMA$(k, k-1)$ processes as $\mathbf{R}_k$, where each element of $\mathbf{R}_k$ is of the form

$$\mathbf{R}_k^{(m,n)} := \left\{ c^{(m,n)} \frac{\prod_{j=1}^{k-1}(1 - b_j^{(m,n)} z)}{\prod_{i=1}^{k}(1 - a_i^{(m,n)} z)} : \ a_i^{(m,n)}, b_j^{(m,n)}, c^{(m,n)} \in \mathbb{C}, \ |a_i^{(m,n)}| < 1, \ \forall i, j \right\}$$

for $m = 1, 2, \ldots, n_x$ and $n = 1, 2, \ldots, n_\epsilon$. We will first show that $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k$ is dense in the normed vector space $\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$ and then extend the result to $\bigcup_{k\in\mathbb{N}} \mathbf{Q}_k$.

By definition, $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k$ is dense in the normed vector space $\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$ if for any $\epsilon > 0$ and any matrix $K(z) \in \mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$, there exists an element $J(z) \in \bigcup_{k\in\mathbb{N}} \mathbf{R}_k$ such that

$$\|K(z) - J(z)\|_{\mathbf{H}^2_{n_x\times n_\epsilon}} = \left(\frac{1}{2\pi i} \oint_{\mathbb{T}} tr \left\{ [K(z) - J(z)] \, [K(z) - J(z)]^* \right\} \frac{dz}{z} \right)^{1/2} \leqslant \epsilon$$

where $*$ denotes conjugate transpose, and $tr(\cdot)$ is the matrix trace operator. We proceed the proof in three steps.

Step 1: We first show that each matrix element, $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k^{(m,n)}$, is dense in $\mathbf{H}^2(\mathbb{D})$. Fix an element (m, n), the proof of this step is constructive. Consider a sequence of complex numbers $\{\theta_k\}_{k=0}^{\infty}$ on the open unit disk $\mathbb{D}$ such that $\lim_{k\to\infty} |\theta_k| = 0$. It is immediate that

$$\sum_{k=0}^{\infty} (1 - |\theta_k|) = \infty \tag{1}$$

Then we apply the Gram-Schmidt procedure to construct an orthonormal basis in the Hilbert space $\mathbf{H}^2(\mathbb{D})$. In particular, consider a set of functions $\{\mathcal{H}_k(z)\}_{k\in\mathbb{N}}$ given by

$$\mathcal{H}_k(z) = \frac{1}{1 - \theta_k z}, \qquad k \in \mathbb{N}$$

The first element can be normalized as

$$\mathcal{B}_0(z) = \frac{\mathcal{H}_0(z)}{\|\mathcal{H}_0(z)\|} = \frac{\sqrt{1-|\theta_0|^2}}{1-\theta_0 z}$$

which simply characterizes an AR(1) process with unit variance (or norm). Next, we recursively define

$$\mathcal{W}_k(z) = \mathcal{H}_k(z) - \sum_{h=0}^{k-1} < \mathcal{H}_k(z), \mathcal{B}_h(z) > \mathcal{B}_h(z), \qquad \mathcal{B}_k(z) = \frac{\mathcal{W}_k(z)}{\|\mathcal{W}_k(z)\|}$$

where the inner product is defined as

$$< \mathcal{H}_k(z), \mathcal{B}_h(z) >= \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathcal{H}_k(e^{-i\omega})\overline{\mathcal{B}_h(e^{-i\omega})}d\omega$$

Note that the choice of $\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}$ is not unique, provided the underlying ARMA process is not too persistent (i.e., θ_k is not too close to unity).

The above recursion yields the following formula for $\{\mathcal{B}_n(z)\}_{n\in\mathbb{N}}$

$$\mathcal{B}_k(z) = \left(\frac{\sqrt{1-|\theta_k|^2}}{1-\theta_k z}\right)\prod_{h=0}^{k-1}\frac{z-\theta_h}{1-\theta_h z}, \qquad k \in \mathbb{N} \tag{2}$$

By inspection, it is easy to see that $\mathcal{B}_k(z) \in \mathbf{R}_k^{(m.n)}$ for all $k \in \mathbb{N}$. Thus,

$$\operatorname{span}\left(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}\right) \subseteq \bigcup_{k\in\mathbb{N}} \mathbf{R}_k^{(m,n)} \tag{3}$$

as $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k^{(m.n)}$ is a linear subspace of $\mathbf{H}^2(\mathbb{D})$ that is closed under finitely many linear combinations.

Therefore, (3) implies that it suffices to show $\operatorname{span}\left(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}\right)$ is dense in $\mathbf{H}^2(\mathbb{D})$. Basic Hilbert space theory ensures that $\operatorname{span}\left(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}\right)$ is dense if and only if there is no function $f(z) \neq 0$ in $\mathbf{H}^2(\mathbb{D})$ such that $< f(z), g(z) >= 0$ for all $g(z) \in \operatorname{span}\left(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}\right)$. We prove this statement by contradiction. Suppose there exists $f(z) \neq 0$ such that it is orthogonal to every element in $\operatorname{span}\left(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}}\right)$. Then we can have

$$< f(z), \mathcal{B}_0(z) >= \frac{1}{2\pi i}\oint_{\mathbb{T}} f(z)\overline{\mathcal{B}_0(z)}\frac{dz}{z} = \frac{1}{2\pi i}\oint_{\mathbb{T}} f(z)\left(\sqrt{1-|\theta_0|^2}\right)\frac{z}{z-\overline{\theta}_0}\frac{dz}{z} = 0 \tag{4}$$

By Morera's Theorem and Cauchy Integral Theorem, (4) holds if and only if $f(z)$ has a zero at $z = \overline{\theta}_0$. Continuing this argument, we know that $f(z)$ has zeros at a sequence of points

$\{\overline{\theta}_k\}_{k=0}^{\infty}$ inside the unit circle. Since $f(z)$ is analytic inside the unit circle, by Theorem 2.3 and its corollary of Duren (2000), we have $f(z) \in \mathbf{H}^2(\mathbb{D})$ if and only if

$$\sum_{k=0}^{\infty} \left(1 - |\overline{\theta}_k|\right) = \sum_{k=0}^{\infty} (1 - |\theta_k|) < \infty$$

which leads to an immediate contradiction to (1). Therefore, span $(\{\mathcal{B}_k(z)\}_{k\in\mathbb{N}})$ and hence $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k^{(m,n)}$ are dense in $\mathbf{H}^2(\mathbb{D})$. Since our choice of (m, n) is arbitrary, the argument extends to all elements of the matrices in the set $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k$.

Step 2: Given step 1, the second step is straightforward. Fix a $K(z) \in \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ and $\epsilon > 0$. For each $m = 1, 2, ...n_x$ and $n = 1, 2, ...n_\epsilon$, pick a function $g^{(m,n)}(z) \in \bigcup_{k\in\mathbb{N}} \mathbf{R}_k^{(m,n)}$ such that

$$\left\|g^{(m,n)}(z) - K^{(m,n)}(z)\right\|_{\mathbf{H}^2} \leqslant \frac{\epsilon}{n_x n_\epsilon}.$$

This can be done by the denseness proved in step 1. Define $J(z)$ such that $J^{(m,n)}(z) = g^{(m.n)}(z), \forall m, n$. Then it follows that

$$\begin{aligned}
&\left(\frac{1}{2\pi i}\oint_{\mathbb{T}} tr\left\{[K(z) - J(z)]\,[K(z) - J(z)]^*\right\}\frac{dz}{z}\right)^{1/2} = \left(\frac{1}{2\pi i}\oint_{\mathbb{T}} \sum_{m=1}^{n_x}\sum_{n=1}^{n_\epsilon}\left|J^{(m,n)}(z) - K^{(m,n)}(z)\right|^2 \frac{dz}{z}\right)^{1/2} \\
&= \left(\sum_{m=1}^{n_x}\sum_{n=1}^{n_\epsilon}\frac{1}{2\pi i}\oint_{\mathbb{T}}\left|J^{(m,n)}(z) - K^{(m,n)}(z)\right|^2\frac{dz}{z}\right)^{1/2} = \left(\sum_{m=1}^{n_x}\sum_{n=1}^{n_\epsilon}\left\|J^{(m,n)}(z) - K^{(m,n)}(z)\right\|^2_{\mathbf{H}^2}\right)^{1/2} \\
&\leqslant \sum_{m=1}^{n_x}\sum_{n=1}^{n_\epsilon}\sqrt{\left\|J^{(m,n)}(z) - K^{(m,n)}(z)\right\|^2_{\mathbf{H}^2}} \\
&\leqslant \frac{\epsilon}{n_x n_\epsilon}(n_x n_\epsilon) \\
&= \epsilon
\end{aligned}$$

where the first inequality comes from the classical arithmetic inequality.

Step 3: By definition, $\bigcup_{k\in\mathbb{N}} \mathbf{R}_k \subset \bigcup_{p,q\,\in\mathbb{N}} \mathbf{Q}_{(p,q)}$. Therefore, $\bigcup_{p,q\,\in\mathbb{N}} \mathbf{Q}_{(p,q)}$ is dense in $\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$.[1] The proof is then complete.

Proof of Theorem 3.2

We break the proof of the theorem in three parts.

[1] In fact, any VARMA(p, q) process can be written as a VARMA$(k, k-1)$ process by setting the appropriate coefficient matrices to zero. The refinement of the basis functions to VARMA$(k, k-1)$ processes is needed for models whose solution forms are restricted in order to integrate our APFI method with other algorithms. See the rational inattention example of Miao, Wu and Young (2021a).

Part 1: First, we notice that $\mathbb{D}$ is open, connected sets in the complex domain such that $\mathbb{U} = (-1, 1) \subset \mathbb{D}$ is open. By definition, $\Psi^y(z)$ is analytic in $\mathbb{D}$; therefore, $\Psi^y(z)$ is the unique analytic continuation of $\Psi^y(z), z \in \mathbb{U}$ to the entire open unit disk.

Next, we rewrite (3.12) as

$$A_0\Psi^y(z) = \mathcal{T}\left(\Psi^y(z)\right), \qquad z \in \mathbb{U} = (-1, 1), \tag{5}$$

where $\mathcal{T} \equiv -\sum_{k=1}^{l} A_k z^k \Psi^y(z) - \sum_{k=0}^{h} B_k F_k(\Psi^y(z))\Psi^{\Omega}(z)$. The first part of $\mathcal{T}$ is analytic and $\left\|z^k\Psi^y(z)\right\|_{\mathbf{H}^2} = \|\Psi^y(z)\|_{\mathbf{H}^2}$, since z^k is merely the shift operator. The Wiener-Hopf operator is defined as

$$F_k(\Psi^y(z)) = \left[z^{-k}\Psi^y(z)\Sigma_\epsilon\Gamma^{\Omega}\left(z^{-1}\right)'\left(\widetilde{\Gamma}^{\Omega}\left(z^{-1}\right)'\right)^{-1}\right]_+ \Sigma_u^{-1}\widetilde{\Gamma}^{\Omega}(z)^{-1}$$

It is easy to show that the annihilation operator $[\cdot]_+$ is linear and the resulting function is analytic in $\mathbb{D}$ (see e.g. Hansen and Sargent (1980) and the proof of the next theorem). The fundamental spectral factor $\widetilde{\Gamma}^{\Omega}(z)^{-1}$ is also analytic in $\mathbb{D}$ by invertibility. Hence, $F_k(\Psi^y(z))$ and $\mathcal{T}\left(\Psi^y(z)\right)$ are analytic functions in the entire open unit disk since analyticity is preserved under sum and product.

Since both the LHS and RHS of (5) are analytic functions in $\mathbb{D}$ and (5) holds in the open subset $\mathbb{U} \subset \mathbb{D}$, we have

$$A_0\Psi^y(z) = \mathcal{T}\left(\Psi^y(z)\right), \qquad z \in \mathbb{D}, \tag{6}$$

by the uniqueness of analytic continuation (Rudin (1987), Corollary of Theorem 10.18) . Moreover, $A_0\Psi^y(z)$ and $\mathcal{T}\left(\Psi^y(z)\right)$ have the same unique Laurent series expansion in the annulus $0 \leqslant |z| < 1$ (i.e., $\mathbb{D}$). Since $\Psi^y(z) \in \mathbf{H}^2_{(2n_x+n_s)\times n_\epsilon}(\mathbb{D})$, by the Riesze-Fischner theorem the Laurent series are square–summable (Theorem 17.12 of Rudin (1987)). Therefore, the MA(∞) representation $\Psi^x(\mathrm{L}) = \sum_{n=0}^{\infty} \Psi_n^x \mathrm{L}^n$ is a covariance-stationary equilibrium for model (3.1).

Part 2: The statement is standard implication of the uniqueness of the analytic continuation and hence proof is omitted.

Part 3: The proof on the convergence criteria is much more involved since analytic continuation does not extend to limit in general. Therefore we adopt a different approach that requires some results in complex analysis. Let $(\mathbb{C}^{n_x\times n_\epsilon}, ||\cdot||_{hs})$ denote the set of $n_x \times n_\epsilon$ dimensional complex matrices equipped with the Hilbert-Schmidt matrix norm, which is the underlying field for matrix functions $\{\Gamma_n^x(z)\}_{n\in\mathbb{N}}$. Since every finite dimensional normed vector space is Banach and satisfies the Heine-Borel theorem, it induces a complete metric space denoted by ($\mathbb{C}^{g\times k}$, d_{hs}).[2] Clearly, $||\cdot||_{hs}$ is a matrix generalization of the univariate modulus $|\cdot|$.

[2]We pick the (Euclidean) H-S norm for convenience as all norms on $\mathbb{C}^{g\times k}$ are equivalent.

Step 1: First, note the each element in sequence $\{\Gamma_n^x(z)\}_{n\in\mathbb{N}} \in \mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$ is rational. Therefore, by Theorem S1.4, they have no poles (and are analytic) on the closed unit disk $\mathbb{T}\bigcup\mathbb{D}$. Moreover, they are bounded analytic functions in the sense

$$\sup_{z\in\mathbb{D}} \|\Gamma_n^x(z)\|_{hs} \leqslant \sup_{|z|=1} \|\Gamma_n^x(z)\|_{hs} = M_n < \infty, \quad \forall n \in \mathbb{N}$$

for some $M_n > 0$. The first inequality follows from the Maximum Modulus Principle. Therefore, $\{\Gamma_n^x(z)\}_{n\in\mathbb{N}} \in \mathbf{H}^\infty_{n_x\times n_\epsilon}(\mathbb{D})$. Define $M = \sup_{n\in\mathbb{N}}\{M_n\} < \infty$. Then it is clear that the sequence $\{\Gamma_n^x(z)\}_{n\in\mathbb{N}}$ is uniformly bounded on every compact subsets of $\mathbb{D}$, i.e., for every compact set $G \subset \mathbb{D}$ and for all $\Gamma_n^x(z)$ and $z \in G$, we have

$$\|\Gamma_n^x(z)\|_{hs} \leqslant M, \quad n \in \mathbb{N}$$

Step 2: Next, we state the following lemma which is crucial for our proof.

Lemma 1 (**Vitali**). *Let f_n be a sequence of analytic functions on a domain D that is uniformly bounded on each compact subset of D. Then the functions f_n converge to f uniformly on compact subsets of D if and only if there is a set of points A, such that A has a point of accumulation in D and f_n converge pointwise on A.*

The proof of this theorem uses the celebrated Montel's theorem and can be found in Beliaev (2019), Theorem 2.11.[3] Now let $D = \mathbb{D}$ and $A = \mathbb{U}$, it is clear that $\{\Gamma_n^x(z)\}_{n\in\mathbb{N}}$ converges uniformly to $\Gamma^x(z)$ on each compact subset of $\mathbb{D}$.

Step 3: We prove the statement $\lim_{n\to\infty} \|\Gamma^x(z) - \Gamma_n^x(z)\|_{\mathbf{H}^2} = 0$. Consider a family of compact sets $\{\bar{D}_r\}_{0\leqslant r<1}$, where $\bar{D}_r$ denotes the closed disk centered at the origin, with radius $0 \leqslant r < 1$. Fix a r and $\epsilon > 0$. Then we can pick $n_0 \in \mathbb{N}$ such that for all $n \geqslant n_0$,

$$\sup_{z\in\bar{D}_r} \|\Gamma^x(z) - \Gamma_n^x(z)\|_{ds} < \epsilon$$

by the uniform convergence property. Now pick $m, n \geqslant n_0$, and define a family of parameterized integral operator $\Delta(\Gamma_n^x(z) - \Gamma_m^x(z), r)$ indexed by r such that

$$\begin{aligned}
\Delta(\Gamma_n^x(z) - \Gamma_m^x(z), r) &= \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left\{\left[\Gamma_n^x(re^{it}) - \Gamma_m^x(re^{it})\right]\left[\Gamma_n^x(re^{it}) - \Gamma_m^x(re^{it})\right]^*\right\}dt\right\}^{\frac{1}{2}} \\
&\leqslant \left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\sup_{z\in\bar{D}_r} \|\Gamma_m^x(z) - \Gamma_n^x(z)\|_{ds}\right|^2 dt\right\}^{\frac{1}{2}} = \sup_{z\in\bar{D}_r} \|\Gamma_m^x(z) - \Gamma_n^x(z)\|_{ds}
\end{aligned}$$

[3] For more background details on the Montel space of analytic functions, we refer readers to Chapter VII of Conway (1978) and Chapter 2 of Beliaev (2019).

where the first inequality follows form the monotonicity of the Lebesgue integrals and the fact that $\sup_{z\in\bar{D}_r}\|\Gamma^x_m(z)-\Gamma^x_n(z)\|_{hs}$ is a constant function. Using the triangle inequality and the property of supremum

$$\begin{aligned}\Delta(\Gamma^x_n(z)-\Gamma^x_m(z),r) &\leqslant \sup_{z\in\bar{D}_r}\|\Gamma^x_m(z)-\Gamma^x_n(z)\|_{ds}\\ &= \sup_{z\in\bar{D}_r}\|\Gamma^x_m(z)-\Gamma^x(z)+\Gamma^x(z)-\Gamma^x_n(z)\|_{hs}\\ &\leqslant \sup_{z\in\bar{D}_r}\|\Gamma^x_m(z)-\Gamma^x(z)\|_{hs}+\sup_{z\in\bar{D}_r}\|\Gamma^x(z)-\Gamma^x_n(z)\|_{hs}\end{aligned} \tag{7}$$

Note that (7) holds for all $m,n\geqslant n_0$. Now take the limit $m\to\infty$,

$$\begin{aligned}\lim_{m\to\infty}\Delta(\Gamma^x_n(z)-\Gamma^x_m(z),r) &= \Delta(\Gamma^x_n(z)-\lim_{m\to\infty}\Gamma^x_m(z),r)\\ &\leqslant \lim_{m\to\infty}\sup_{z\in\bar{D}_r}\|\Gamma^x_m(z)-\Gamma^x(z)\|_{hs}+\lim_{m\to\infty}\sup_{z\in\bar{D}_r}\|\Gamma^x(z)-\Gamma^x_n(z)\|_{hs}\end{aligned} \tag{8}$$

where the first equality interchanges the integration with limit under uniform convergence. By Lemma A.1, the first term on the RHS of the inequality vanishes; hence, (8) implies that

$$\Delta(\Gamma^x_n(z)-\Gamma^x(z),r)\leqslant \sup_{z\in\bar{D}_r}\|\Gamma^x(z)-\Gamma^x_n(z)\|_{hs}<\epsilon \tag{9}$$

Since (9) holds for all $0\leqslant r<1$, we take take the radial limit $r\to 1$,

$$\lim_{r\to 1}\Delta(\Gamma^x_n(z)-\Gamma^x(z),r)=\|\Gamma^x_n(z)-\Gamma^x(z)\|_{\mathbf{H}^2}<\epsilon$$

where the first equality follows from Remark 17.8 of Rudin (1987). Now since our choices of n_0 and n are arbitrary, we have proven the statement:

$$\lim_{n\to\infty}\|\Gamma^x(z)-\Gamma^x_n(z)\|_{\mathbf{H}^2}=0$$

By the Cauchy completeness of $\mathbf{H}^2$ space, we know that $\Gamma^x(z)\in\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$. Our proof is now complete.

Proof of Theorem 3.3

First, consider the conditional expectation $\hat{x}_{t+k}=\mathbb{E}[x_{t+k}|\Omega^t]$ in its innovation representation

$$\hat{x}_{t+k}=\sum_{k=0}^{\infty}H_k\,\mathrm{L}^k\,\Omega_t=H(\mathrm{L})\widetilde{\Gamma}^{\Omega}(\mathrm{L})u_t \tag{10}$$

By covariance-stationarity, $H(z)\widetilde{\Gamma}^{\Omega}(z) \in \mathbf{H}^2(\mathbb{D})$. Since $\widetilde{\Gamma}^{\Omega}(z)$ is an outer (i.e., invertible) function, it follows that $H(z)$ is analytic on $\mathbb{D}$ by the properties of analytic functions.[4]

Next, we derive the region of convergence (ROC) for $\Theta(z)$ through the construction of the Wiener-Hopf filter. In particular, we show such ROC contains the unit circle $\mathbb{T}$ under our assumptions. Note that the optimal filter satisfies the following orthogonality condition

$$\mathbb{E}\left[\left(x_{t+k} - \hat{x}_{t+k}\right)\Omega_{t-j}\right] = 0, \qquad \forall k, j \in \mathbb{N} \tag{11}$$

in the inner-product space of random variables. Substitute (10) into (11) to obtain

$$R_{x\Omega}(k+j) = \sum_{i=0}^{\infty} H_i R_{\Omega}(j-i), \qquad \forall k, j \in \mathbb{N} \tag{12}$$

where $R_{x\Omega}(\cdot)$ and $R_{\Omega}(\cdot)$ are cross and auto-covariance functions, respectively. Since $\Gamma^x(z)$ and $\Gamma^{\Omega}(z)$ are rational analytic functions on $\mathbb{T}$ by Theorem S1.4, the implied covariance generating functions

$$S_{x\Omega}(z) \;=\; \Gamma^x(z)\Sigma_\epsilon\Gamma^{\Omega}\left(z^{-1}\right)' = \sum_{k=-\infty}^{\infty} R_{x\Omega}(k) z^k \tag{13}$$

$$S_{\Omega}(z) \;=\; \Gamma^{\Omega}(z)\Sigma_\epsilon\Gamma^{\Omega}\left(z^{-1}\right)' = \sum_{k=-\infty}^{\infty} R_{\Omega}(k) z^k \tag{14}$$

are rational with the property that $\sum_{k=-\infty}^{\infty} |R_{x\Omega}(k)| < \infty$ and $\sum_{k=-\infty}^{\infty} |R_{\Omega}(k)| < \infty$. Geometrically, these functions possess a finite number of poles located inside and outside the unit circle, which are determined by $\Gamma^x(z)$ and $\Gamma^{\Omega}(z)$. Then we can pick a small $\epsilon > 0$ such that within the annulus $\mathbb{A}_1 = \{z : |1-\epsilon| < |z| < 1/|1-\epsilon|\}$, (13) and (14) converge and hence are well-defined. This procedure can simply be done by taking the intersection of the ROCs for $S_{x\Omega}(z)$ and $S_{\Omega}(z)$ that contain the unit circle.

Now define a sequence $\{G_j\}_{j=-\infty}^{\infty}$ by

$$G_j = R_{x\Omega}(k+j) - \sum_{i=0}^{\infty} H_i R_{\Omega}(j-i), \qquad \forall k \in \mathbb{N}, \qquad j \in \mathbb{Z} \tag{15}$$

By (12), $G_j = 0$ for $j \geqslant 0$. We can also extend the sequence $\{H_i\}_{i=0}^{\infty}$ by letting $H_i = 0$ for $i < 0$. Then taking the two-sided z-transform on both sides of (15), we can obtain

$$G(z) = z^{-k} S_{x\Omega}(z) - H(z) S_{\Omega}(z) \tag{16}$$

[4]The Hardy space, however, is not closed under multiplication.

where we have applied the convolution theorem of z-transform. (16) is well-defined in the region $\mathbb{A}_2 = \mathbb{A}_1 \bigcap \mathbb{D} = \{z : |1-\epsilon| < |z| < 1\}$. Using the canonical spectral factorization $S_\Omega(z) = \widetilde{\Gamma}_\Omega(z)\widetilde{\Gamma}_\Omega(z^{-1})'$, we can rewrite (16) as

$$G(z)\left(\widetilde{\Gamma}_\Omega(z^{-1})'\right)^{-1} = z^{-k}S_{x\Omega}(z)\left(\widetilde{\Gamma}_\Omega(z^{-1})'\right)^{-1} - H(z)\widetilde{\Gamma}_\Omega(z)$$

Within the region $\mathbb{A}_2$, $G(z)\left(\widetilde{\Gamma}_\Omega(z^{-1})'\right)^{-1}$ has only negative power terms by the invertibility of the fundamental spectral factor, while $H(z)\widetilde{\Gamma}_\Omega(z)$ has only positive power terms. Therefore, taking the annihilation $[\cdot]_+$ of the above equation yields

$$0 = \left[z^{-k}S_{x\Omega}(z)\left(\widetilde{\Gamma}_\Omega(z^{-1})'\right)^{-1}\right]_+ - H(z)\widetilde{\Gamma}_\Omega(z)$$

which gives the Wiener-Hopf optimal prediction formula

$$H(z) = [\Theta(z)]_+ \widetilde{\Gamma}_\Omega(z)^{-1}$$

By construction $\Theta(z)$ is rational. Moreover, it does not have any pole on the unit circle. In other words, $\Theta(z)$ has the same Laurent series expansion in the region $\mathbb{A}_3 = \{z : |1-\epsilon| < |z| \leqslant 1\}$ as in $\mathbb{A}_2$. Therefore, we can apply the inverse discrete time Fourier transform to compute the coefficients of positive power terms. Finally, since $\Theta(z)$ is a rational analytic function on $\mathbb{T}$, $\Theta(z) \in L^2(\mathbb{T})$ and its inverse Fourier coefficients are absolutely-summable (and hence square-summable). The approximation accuracy follows directly from Theorem S1.3.

Properties of the Baseline Algorithm

Proposition 2 (Boundedness)**.** *Suppose Assumption 4.1 holds for the simplified model system (4.1), where the expectational block is associated with either (i) a particular information set Ω_t (I.0) or (ii) information structure I.1, and the structural innovations are orthonormal, i.e., $\Sigma_\epsilon = I$. Then $\mathcal{A} : \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D}) \mapsto \mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ is a nonlinear, locally bounded operator that maps a bounded subset of $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$ to another bounded subset. If in addition,*

$$\sum_{k=0}^{h} ||A^x(z)^{-1}B_k||_{\mathbf{H}^\infty} = \sum_{k=0}^{h} \sup_{|z|=1} \sigma_{\max}\left(A^x(z)^{-1}B_k\right) < 1 \tag{17}$$

where $\mathbf{H}^\infty$ is defined in the space of (essentially) bounded analytic functions and $\sigma_{\max}$ refers to the largest singular value of the matrix. Then for any given initial conjecture $\Gamma^x_0(z)$, the sequence $\{\Gamma^x_n(z)\}_{n=0}^\infty$ induced by $\mathcal{A}$ is bounded (non-explosive).

The $\mathbf{H}^\infty$ norm is used extensively in the signal processing and control theory literature and can be computed conveniently in MATLAB. The assumption of orthonormality is not restrictive because the fixed point conditions (4.1) and (4.2) can always be rewritten in terms of orthonormal innovations. Specifically, if we perform the (unitary) eigen-decomposition for $\Sigma_\epsilon = UDU'$ with U being unitary and D being diagonal and multiply the structural innovations in $\Gamma^s(z)$, $\Gamma^y(z)$, and $\Gamma^\Omega(z)$ by $M = U\sqrt{D}$, then the fixed point condition can be defined in terms of orthonormal innovations. The solution $\Gamma^x(z)$ to the normalized system is equivalent to the solution with respect to the original structural innovations up to a transformation by M^{-1}.

Proof of Proposition 2

We first consider the case where the expectational block contains a particular information set Ω_t. Expanding (4.2) to obtain

$$\Gamma^x(z) = \mathcal{A}\left(\Gamma^x(z)\right) \equiv -A^x(z)^{-1}A^s(z)\Gamma^s(z) - A^x(z)^{-1}\left(\sum_{k=0}^{h} B_k^x F_k(\Gamma^x(z))\Gamma^\Omega(z) + B_k^s F_k(\Gamma^s(z))\Gamma^\Omega(z)\right) \tag{18}$$

By Assumption 4.1, $A^x(z)^{-1}$ is a rational, bounded analytic function in $\mathbf{H}^\infty_{n_x\times n_x}$. By Conway (1990, p. 28, Theorem 1.5) and Lindquist and Picci (2015, Theorem 4.3.3 (Bochner-Chandrasekharan) and Proposition B.2.4), the left multiplication by $A^x(z)^{-1}$ defines a bounded, linear operator in $\mathbf{H}^2_{n_x\times n_\epsilon}$ with operator norm

$$\left|\left|A^x(z)^{-1}\right|\right|_{op} = \left|\left|A^x(z)^{-1}\right|\right|_{H^\infty} = \sup_{|z|=1} \sigma_{\max}\left(A^x(z)^{-1}\right)$$

Now consider a bounded set in $\mathbf{H}^2_{n_x\times n_\epsilon}(\mathbb{D})$ and an element $\Gamma^x(z)$ in it. It follows that

$$\begin{aligned}\left|\left|\mathcal{A}\left(\Gamma^x(z)\right)\right|\right|_{\mathbf{H}^2} \leqslant{}& \left|\left|A^x(z)^{-1}A^s(z)\Gamma^s(z)\right|\right|_{\mathbf{H}^2} + \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^x\right|\right|_{op}\left|\left|F_k(\Gamma^x(z))\Gamma^\Omega(z)\right|\right|_{\mathbf{H}^2}\\ &+ \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^s\right|\right|_{op}\left|\left|F_k(\Gamma^s(z))\Gamma^\Omega(z)\right|\right|_{\mathbf{H}^2}\end{aligned}$$

where we have used the triangular inequality and the bounded linear operators defined by $A^x(z)^{-1}B_k^x$ and $A^x(z)^{-1}B_k^s$, $k = 0, 1, \dots, h$. The first term on the right hand side is purely exogenous and bounded. Without loss of generality, let $N = \left|\left|A^x(z)^{-1}A^s(z)\Gamma^s(z)\right|\right|_{\mathbf{H}^2}$. Define $W\left(\Gamma^x(z)\right) = F_0(\Gamma^x(z))\Gamma^\Omega(z)$, which is associated with the innovation representation for the conditional expectation of the vector-valued process x_t. By the spectral theory of time series, the covariance matrix of $\mathbb{E}_t x_t$ is given by the inverse Fourier transform of the spectral density[5]

$$\Sigma_{Ex} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_{Ex}(\omega)d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega)W^*(\omega)d\omega$$

Taking the trace operator on both sides of the equation leads to the sum of variances

$$tr(\Sigma_{Ex}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(W(\omega)W^*(\omega)\right)d\omega \leqslant tr(\Sigma_x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} tr\left(\Gamma^x(\omega)\Gamma^{x,*}(\omega)\right)d\omega \tag{19}$$

where Σ_x is the covariance matrix of x_t, and the inequality follows from the variance bounds of higher-order expectations. That is, for the entire vector of model variables y_t, we have

$$var(\overline{\mathbb{E}}_t^{(k)}(y_t)) \leqslant var(\overline{\mathbb{E}}_t^{(k-1)}(y_t)) \leqslant \ldots \leqslant var(\overline{\mathbb{E}}_t^{(1)}(y_t)) \leqslant var(\mathbb{E}_t(y_t)) \leqslant var(y_t), \quad \forall k \in \mathbb{N} \tag{20}$$

where $\overline{\mathbb{E}}_t^{(k)}(y_t)$ denotes the k-th order average expectation over the entire economy. $\leqslant$ is defined as entry-wise inequality over vectors. The last inequality follows from the orthogonality condition of individual conditional expectations, and the remaining inequalities follow from the fact that idiosyncratic shocks vanish due to the law of large numbers.

Since x_t is covariance-stationary, the inequality in (19) applies to expectations of future realizations $\{L^{-k}x_t\}_{k=1}^h$. Taking square roots on both sides of the inequality in (19), we obtain the norm inequality: $\left|\left|F_k(\Gamma^x(z))\Gamma^\Omega(z)\right|\right|_{\mathbf{H}^2} \leqslant \left|\left|\Gamma^x(z)\right|\right|_{\mathbf{H}^2}$, $k = 0, 1, \ldots, h$. A similar argument shows that $\left|\left|F_k(\Gamma^s(z))\Gamma^\Omega(z)\right|\right|_{\mathbf{H}^2} \leqslant \left|\left|\Gamma^s(z)\right|\right|_{\mathbf{H}^2}$, $k = 0, 1, \ldots, h$. Therefore,

$$\left|\left|\mathcal{A}\left(\Gamma^x(z)\right)\right|\right|_{\mathbf{H}^2} \leqslant N + \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^x\right|\right|_{op}\left|\left|\Gamma^x(z)\right|\right|_{\mathbf{H}^2} + \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^s\right|\right|_{op}\left|\left|\Gamma^s(z)\right|\right|_{\mathbf{H}^2}$$

is bounded. The last component in the inequality is again exogenous, and we define $R = \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^s\right|\right|_{op}\left|\left|\Gamma^s(z)\right|\right|_{\mathbf{H}^2}$.

Next, we notice that $\left|\left|A^x(z)^{-1}B_k^x\right|\right|_{op} \leqslant ||A^x(z)^{-1}B_k||_{op}$ since the maximum singular value of the sub-matrix is bounded by the original matrix for every $|z| = 1$ and $k = 0, 1, \ldots, h$. Therefore, if (17) holds, let $\alpha \equiv \sum_{k=0}^{h}\left|\left|A^x(z)^{-1}B_k^x\right|\right|_{op} < 1$. Now take an initial conjecture $\Gamma_0^x(z)$. The sequence $\{\Gamma_n^x(z)\}_{n=0}^\infty$ induced by $\mathcal{A}$ admits the bound

$$\left|\left|\Gamma_n^x(z)\right|\right|_{\mathbf{H}^2} = \left|\left|\mathcal{A}\left(\Gamma_{n-1}^x(z)\right)\right|\right|_{\mathbf{H}^2} \leqslant \sum_{k=0}^{n-1}\alpha^k(N+R) + \alpha^n\left|\left|\Gamma_0^x(z)\right|\right|_{\mathbf{H}^2}$$

Since $\alpha \in (0,1)$, $R > 0$, and $N > 0$, the sequence $\{\Gamma_n^x(z)\}_{n=0}^\infty$ is bounded above by $\frac{1}{1-\alpha}(N +$

[5] By Rozanov (1967, Sec 1.9), the processes x_t and $\mathbb{E}_t x_t$ have absolutely continuous spectral measures with well-defined spectral densities.

$R) + \left|\left|\Gamma_0^x(z)\right|\right|_{\mathbf{H}^2}$. Under the more general information structure I.1, elements in $\{B_k\}_{k=0}^h$ is allowed to be associated with expectations conditional on distinct information sets. However, for each information set, the variance bound inequalities (19) and (20) still hold. Therefore, using the representation (3.3) it is easy to show that the results we have proved remain valid. This completes the proof.

Proposition 3 (Contraction Mapping)**.** *Suppose the information structure is exogenous and the assumptions in Proposition 2 hold (in particular, condition (17)). Then $\mathcal{A}$ is a contraction mapping and there exists a unique fixed point for the simplified model system (4.1).*

Proof of Proposition 3

We first consider the case where the expectational block contains only one information set Ω_t. When information is exogenous, by the Wiener-Hopf prediction formula (3.7) and the variance bound inequality (19), $F_k(\Gamma^x(z))\Gamma^\Omega(z)$ defines a linear bounded operator for $k = 0, 1, 2, ...h$. The linearity follows from the fact that the annihilation operator $[\cdot]_+$ is linear and that information is purely exogenous. Next, we show that under (17), $\mathcal{A}$ is a contraction. Take two functions $f(z)$ and $g(z)$ in the space $\mathbf{H}^2_{n_x \times n_\epsilon}(\mathbb{D})$. By (18),

$$\begin{aligned}\mathcal{A}(f-g)(z) &= A^x(z)^{-1}\left(\sum_{k=0}^{h} B_k^x F_k(g(z))\Gamma^\Omega(z) - \sum_{k=0}^{h} B_k^x F_k(f(z))\Gamma^\Omega(z)\right) \\ &= A^x(z)^{-1}\left(\sum_{k=0}^{h} B_k^x F_k(g(z)-f(z))\Gamma^\Omega(z)\right)\end{aligned}$$

where the terms associated with exogenous shocks and their expectations drop out due to exogenous information. The second equality comes from the linearity of the expectation operator. Then it follows that

$$\left|\left|\mathcal{A}(f-g)(z)\right|\right|_{\mathbf{H}^2} \leqslant \sum_{k=0}^{h} \left|\left|A^x(z)^{-1}B_k^x\right|\right|_{op} \left|\left|(f-g)(z)\right|\right|_{\mathbf{H}^2} = \alpha\left|\left|(f-g)(z)\right|\right|_{\mathbf{H}^2}$$

where $\alpha \in (0, 1)$ is defined in the proof of Proposition 2. Therefore, $\mathcal{A}$ is a contraction and there is a unique fixed point for (4.1). Using the representation (3.3), the proof for more general types of information structure is similar and hence omitted.