Macroeconomic Models with Incomplete Information and Endogenous Signals *

Jonathan J Adams[†]

December 24, 2024

Link to Most Current Version

Abstract

This paper characterizes a general class of macroeconomic models with incomplete information, which feature endogenous signal processes. These types of models are not always well-behaved, possibly featuring many equilibria, and solution algorithms may not converge to a fixed point. I introduce an *Information Feedback Regularity* condition to discipline these models. The regularity condition is necessary for a stable fixed point to exist. Stable fixed points have nice properties: finite-dimensional fixed points approximate infinite-dimensional stable equilibria arbitrarily well, and iterative algorithms will converge to them. Most importantly, I prove a global uniqueness theorem: if an equilibrium fixed point is stable, then it is the unique stable equilibrium. Next I derive a sufficient condition; if the signal process includes enough idiosyncratic noise, then all fixed points must be stable, guaranteeing global uniqueness. I study the conditions and equilibrium properties in a number of example applications. Finally, I introduce an algorithm to solve the general model, and provide resources to compute it.

JEL-Codes: D84, E22, E32, C62, C63
Keywords: Endogenous Signals, Incomplete Information, Dispersed Information, Heterogeneous Beliefs, Equilibrium Properties

---

*I am grateful for helpful comments from Chunrong Ai, Ryan Chahrour, Thorsten Drautzburg, Igor Livshits, Kristoffer Nimark, Richard Romano, David Sappington, Todd Walker, seminar participants at the University of Florida, University of Connecticut, Federal Reserve Bank of Philadelphia, and the Summer Meetings of the European Econometric Society. I especially appreciate comments on some of the operator theory from Michael Jury and James Pascoe. This paper previously circulated under the title "Rational Expectations with Endogenous Information." Replication code and resources for applying the solution algorithm to other models are available on my website. All errors are my own.

[†]University of Florida. Website: www.jonathanjadams.com Email: adamsjonathan@ufl.edu

1 Introduction

This paper studies macroeconomic models with dispersed information and endogenous signals.[1] When signals are exogenous, models are well understood: there are conditions that ensure equilibrium existence and uniqueness.[2] However, modelers may prefer signals to be endogenous, so that agents can learn from realistic sources, such as employment or interest rates. When a model features endogenous signals, when will its equilibrium be unique? This is an open question. This paper makes progress towards answering it.

The first theoretical contribution is a regularity condition. When signals are endogenous, agents' choices depend on their information sets, which depend on agents' choices, which depend on their information sets, and so forth. Is this feedback explosive or well-behaved? The *Information Feedback Regularity* condition addresses this question by quantifying this explosiveness.

Checking if a model satisfies Information Feedback Regularity is easy and useful. Crucially, the regularity condition is necessary for a model to have a *signal-stable* equilibrium. Signal-stable equilibria have many nice properties: they are robust to small deviations in the signal process, they are locally unique, and an iterative algorithm is guaranteed to converge to the solution given a good initial guess. Moreover, if they are infinite dimensional, they can be approximated arbitrarily well by finite-dimensional signal-stable solutions. But perhaps the most valuable property of signal-stable equilibria is that their global uniqueness is describable.

When is an equilibrium globally unique? For these types of models, the existing literature cannot say, except in specialized settings.[3] It is extremely challenging to characterize uniqueness in general because solutions are often infinite dimensional and feedbacks are typically non-contractive and nonlinear. However, I demonstrate that it is possible to describe uniqueness within the class of signal-stable equilibria. And again, the regularity condition is crucial. If Information Feedback Regularity holds, then any signal-stable equilibrium must be the globally unique signal-stable equilibrium.

Is it possible to guarantee global uniqueness? Yes, if a model satisfies the *Sufficient Idiosyncrasy Condition*. The condition is satisfied when idiosyncratic noise is

[1] "Endogenous" signals or information has different meanings in different literatures. In this context it is Huo and Takayama (2015)'s definition: the endogeneity refers to agents' observation of noisy signals containing endogenous variables. This contrasts with the large literature of endogenous information acquisition, where agents choose to utilize a subset of available information, such as in the rational inattention literature following Sims (2003).

[2] See for example Han, Tan, and Wu (2022), Huo and Takayama (2023), or Theorem 1 in this paper.

[3] In dynamic models without endogenous state variables, uniqueness or multiplicity can sometimes be characterized. For example, some asset pricing models with endogenous information have demonstrable multiplicity, as in Angeletos and Werning (2006), Hellwig, Mukherji, and Tsyvinski (2006), or Angeletos, Hellwig, and Pavan (2007). In others such as Grossman (1976), the equilibrium is unique.

sufficiently large; this dampens the feedback from aggregate variables to expectations back to aggregate variables. The sufficient condition implies the necessary condition; Information Feedback Regularity must hold as a consequence. The condition guarantees uniqueness by ensuring that all fixed points are signal-stable. Sufficient Idiosyncrasy may not hold in all models; in particular it requires that the model features as many idiosyncratic shocks as signals. But as with the regularity condition, the sufficient condition can be easily checked from the parameterization of a model without solving it.

I derive these results for a general class of macroeconomic models that can include endogenous state variables such as capital.[4] I show how to rewrite these types of models as a nonlinear *signal operator* that maps signal processes to signal processes. A solution to the model is a fixed point of the signal operator; thus the theoretical results derive properties of this operator. For example, the Information Feedback Regularity condition controls how explosive the signal operator is in certain directions. When a fixed point is signal-stable, repeated application of the signal operator will converge to it. I refer to this solution algorithm as *Signal Operator Iteration*. But, most theoretical results in this paper can be applied without using this particular algorithm.[5]

To demonstrate how to apply these findings, I consider a number of simple examples drawn from the literature. It is straightforward to represent a variety of common model structures in the general form described in this paper. From there, the regularity condition is easily determined by calculating the norm of a block Toeplitz matrix. In some cases, this can be done analytically; otherwise I provide code to do so numerically. Then I show that choosing model parameterizations to satisfy the condition is helpful for objectives such as ensuring fixed point existence, ensuring uniqueness, selecting among multiple equilibria on the basis of stability, or understanding numerical non-convergence.

[4]Endogenous states introduce additional challenges, so very few publications study such models without additional assumptions that reveal the true state. The earliest example is Graham and Wright (2010), who solve a version of the Neoclassical growth model with dispersed information. Their model features two signals and two shocks, but there are confounding dynamics so that the aggregate shock is not perfectly invertible from the aggregate signal. A recent example is Adams (2023) which studies an optimal policy when dispersed information amplifies the macroeconomic effects of noise shocks.

[5]The literature has several existing methods to solve models with endogenous information. Past shocks can be revealed to agents so that the information problem remains static, as in (Lucas, 1972), or if there are as many shocks and signals, Blaschke root-flipping can be used to solve a model analytically (e.g. Kasa (2000), Acharya (2013), or Rondina and Walker (2015)). But if shocks are never revealed, solution is more challenging. Nimark (2017) uses an iterative algorithm to calculate higher order expectations in a general asset pricing model with endogenous information; this algorithm can be applied to more general settings without endogenous states, as Nimark (2008) and Melosi (2016) do in New Keynesian models. When a model features endogenous states, another option is Han, Tan, and Wu (2022), who improve upon this paper's methodology by approximating signals with a finite ARMA at each iteration.

The strategy for the remainder of the paper is outlined as follows. In Section 2 I define a general linear rational expectations model with incomplete information, and I derive agents' optimal policy function in terms of the signals they observe, and define the signal operator. In Section 3 I introduce the Information Feedback Regularity condition, describe Signal Operator Iteration, and prove various properties of the operator, including the existence and uniqueness theorems. Section 4 introduces and demonstrates the sufficient condition. Section 5 explores the simple examples applying the method and drawing conclusions from the regularity condition. Section 6 concludes.

2 The General Macroeconomic Model

In this section I describe a general macroeconomic model with incomplete information. I describe the macroeconomic structure, derive agents' optimal policy function, and characterize how the endogenous signal process is determined.

Consider a stationary linear macroeconomic model of the following form.[6] The equilibrium conditions for agent i at time t are:

$$0 = E_{i,t}[B_{X0}X_{i,t} + B_{X1}X_{i,t+1} + B_{A0}A_{i,t} + B_{A1}A_{i,t+1}] \tag{1}$$

$X_{i,t}$ is an $n \times 1$ vector of endogenous variables. $n = n_C + n_S$ where n_C is the number of control variables which are chosen at time t, while n_S is the number of state variables which are chosen at time $t-1$. Assume that $X_{i,t}$ is ordered so that the control variables appear first. The $m_A \times 1$ vector $A_{i,t}$ contains the information observed by agents: m_A linearly independent time series. This may include exogenous variables such as economic shocks and signals, and it may include variables that agent i takes as exogenous, but are endogenous in equilibrium, such as an economy-wide interest rate or price level. When agents form expectations, their information set is the history of the $A_{i,t}$ vectors:

$$E_{i,t}[\cdot] \equiv E[\cdot|\{A_{i,t-j}\}_{j=0}^{\infty}]$$

The matrices $\{B_{X0}, B_{X1}, B_{A0}, B_{A1}\}$ contain coefficients encoding the n equilibrium conditions of the model that determine agent i's choice of the endogenous variables in $X_{i,t}$. Additional equations that determine how $A_{i,t}$ is determined are incorporated when information is endogenized in Section 2.2.1.

[6] This general form encompasses a broad class of DSGE models, and is solved without information frictions in Uhlig (1995), among many others. This structure also nests many popular types of information frictions, including models that do not require forecasting, such as the beauty contests in Section 5.1. The main limitations of the information structure in this paper are that: (1) it does not allow for a discrete number of information sets (Han, Tan, and Wu (2022) introduce a method which can solve such problems) and (2) it does not allow agents' equilibrium conditions to be affected by shocks or prices that do not enter their information set (this excludes models that assume agents only forecast using e.g. a subset of prices).

2.1 The Policy Function

I will derive agents' optimal choices as a policy function where the input is their information set.

A linear solution to the model is policy that expresses $X_{i,t}$ as a function of variables $A_{i,t-k}$ for $k \geq 0$ such that (1) holds with equality for all t. This is not necessarily a recursive policy function; it may depend on the entire history of $A_{i,k}$. Specifically, define policy functions to be linear in the history of white noise forecast errors $W_{i,t}$. With this basis, $A_{i,t}$ is given by

$$A_{i,t} = A(L)W_{i,t} \equiv \sum_{j=0}^{\infty} A_j L^j W_{i,t} \tag{2}$$

When the lag operator polynomial $A(L)$ is invertible, this corresponds to the Wold representation. The policy function can be expressed as a lag operator polynomial:

$$X_{i,t} = X(L)W_{i,t} \equiv \sum_{j=0}^{\infty} X_j L^j W_{i,t} \tag{3}$$

Expressing policy functions in terms of information is convenient because forecasting is straightforward: $E_{i,t}[W_{i,t+k}] = 0$ for all $k > 0$.[7] Frequently the policy function is expressed in term of the history of signals, and this form is easily recovered by inverting $A(L)$:

$$X_{i,t} = X(L)W_{i,t} = X(L)A(L)^{-1}A_{i,t} \tag{4}$$

where $\{X_j\}_{j=0}^{\infty}$ are $n \times m_A$ matrices. When expressed in terms of the innovations $W_{i,t}$, the equilibrium condition (1) becomes

$$0 = B_{X0}X(L)W_{i,t} + [B_{X1}L^{-1}X(L)]_+W_{i,t} + B_{A0}A(L)W_{i,t} + [B_{A1}L^{-1}A(L)]_+W_{i,t} \tag{5}$$

where $[\cdot]_+$ is the annihilation operator, which annihilates negative powers of L. I assume that agents forecast linearly, which is optimal when shocks are normal. Then, equation (5) follows from equation (1).

The equilibrium policy function can be expressed as a linear function of the forecast errors $W_{i,t}$. Before deriving the formula, some notation must be defined.

The generalized Schur decomposition of the coefficient matrices is denoted by

$$B_{X0} = QT_0Z \qquad B_{X1} = QT_1Z$$

[7] This is the Wiener-Kolmogorov prediction formula. See Hansen and Sargent (1981) for a description in the context of rational expectations models. The Wiener filter in this case is used for characterizing expectations in lieu of the Kalman filter is which more common in the literature; the Kalman filter is less convenient in this situation because time is infinite, and when endogenous information is introduced, the state space becomes infinite as well.

where Q and Z are unitary, T_0 and T_1 are upper triangular, and the diagonal of T_0 is arranged so that the generalized eigenvalues are ordered with increasing magnitudes. Partition the matrices into blocks, separating the first n_S dimensions from the remaining n_C dimensions. Denote the partitions as:

$$T_0 = \begin{pmatrix} T_{0,SS} & T_{0,SC} \\ 0 & T_{0,CC} \end{pmatrix} \qquad T_1 = \begin{pmatrix} T_{1,SS} & T_{1,SC} \\ 0 & T_{1,CC} \end{pmatrix} \qquad Z = \begin{pmatrix} Z_{SS} & Z_{SC} \\ Z_{CS} & Z_{CC} \end{pmatrix}$$

I make two regularity assumptions about the model, following Klein (2000). Blanchard and Kahn (1980) make similar assumptions in a less general setting.

1. Z_{CC} is invertible.

2. B_{X0} and B_{X1} have no undefined or unit generalized eigenvalues.[8]

Define the polynomials $\Xi(L)$ and $\Theta(L)$ by

$$\Xi(L) \equiv \begin{pmatrix} 0 & -L^{-1}B_C(L)^{-1}T_{0,CC}^{-1} \\ I & 0 \\ 0 & -B_C(L)^{-1}T_{0,CC}^{-1} \end{pmatrix} Q^* \tag{6}$$

$$\Theta(L) \equiv$$
$$Z^* \begin{pmatrix} -B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} + T_{1,SS}^{-1}T_{1,SC}\right)L & -B_S(L)^{-1}T_{1,SS}^{-1}L & B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} - T_{1,SS}^{-1}T_{0,SC}L\right) \\ 0 & 0 & I \end{pmatrix} \tag{7}$$

where $B_S(L) \equiv (I + T_{1,SS}^{-1}T_{0,SS}L)$ and $B_C(L) \equiv (I + T_{0,CC}^{-1}T_{1,CC}L^{-1})$.

Theorem 1 *If B_{X0} and B_{X1} have exactly n_C generalized eigenvalues outside the unit circle, then the unique policy function is given by*

$$X(L) = \Theta(L)\left[\Xi(L)\left(B_{A1}L^{-1} + B_{A0}\right)A(L)\right]_+$$

Proof: Appendix A.1.

The purpose of Theorem 1 is to express the policy function in a way that can be easily applied to the endogenous information case in Section 2.2. The requirement that n_C eigenvalues are outside the unit circle is not novel; it is equivalent to the Klein (2000) generalization of the Blanchard and Kahn (1980) condition that there

[8]The ith generalized eigenvalue of the model is the ratio of diagonal elements $T_{0,i,i}/T_{1,i,i}$. If $T_{1,i,i}$ is zero while $T_{0,i,i}$ is nonzero, the generalized eigenvalue is said to be infinite. If both are zero, then the generalized eigenvalue is said to be undefined.

must be as many unstable eigenvalues as there are contemporaneous jump variables for the equilibrium to be uniquely determined.[9]

The advantage of expressing the policy function this way is that it provides a single linear operator that maps agents' information (encoded in $A(L)$) to their actions (encoded in $X(L)$). This linearity is valuable for proving properties about the general equilibrium in Section 3 such as existence and stability. Adding and multiplying lag operator polynomials are linear operations, as is applying the annihilation operator.

2.2 Endogenous Information

This section details how the information process is formed, how it depends on endogenous decisions, and the fixed point equation that it must satisfy in equilibrium.

In this section, I assume that the conditions for Theorem 1 are satisfied, so that given an information process, agents have a unique policy function.

2.2.1 Endogenous Information Formation

When information is endogenous, the signals $A_{i,t}$ are jointly determined in equilibrium with the rest of the model. For $A_{i,t}$ to be endogenous, the model requires an additional equilibrium condition. This is the *fixed point equation*: the signal dynamics must be consistent with the dynamics implied by the other endogenous variables. I proceed by outlining a general framework for how endogenous information is formed, characterize it in terms of lag operator polynomials, and then define the fixed points.

Suppose the signals $A_{i,t}$ observed by agent i are a sum of exogenous signals $S_{X,i,t}$ and endogenous signals $S_{N,i,t}$:

$$A_{i,t} = S_{X,i,t} + S_{N,i,t} \tag{8}$$

where all of these signals are $m_A \times 1$ vectors. These signals can be expressed as lag operator polynomials times the white noise process of fundamental exogenous shocks, $\varepsilon_{i,t}$, which has dimensionality $m_\varepsilon \geq m_A$ and (without loss of generality) unit variance:

$$S_{X,i,t} = S_X(L)\varepsilon_{i,t} \qquad S_{N,i,t} = S_N(L)\varepsilon_{i,t} \tag{9}$$

The causal square-summable polynomial $S_X(L)$ is a primitive of the model. But the polynomial $S_N(L)$ depends on equilibrium behavior and aggregation. Define the sum of the two polynomials as

$$A_{i,t} = S(L)\varepsilon_{i,t} \equiv S_X(L)\varepsilon_{i,t} + S_N(L)\varepsilon_{i,t} \tag{10}$$

[9] Huo and Takayama (2023) impose a similar eigenvalue condition to ensure equilibrium existence and uniqueness in general models with exogenous signal processes. They go further and find additional regularity conditions such that the eigenvalue condition is not only sufficient but also necessary.

Endogenous signals are determined by macroeconomic aggregates. This assumes that the actions of atomistic agent i do not affect the information of any agent beyond their effect on the aggregate economy. The square-summable polynomial $G(L)$ encodes exactly how aggregate variables affect the endogenous signal. For example, it may include aggregate resource constraints or adding up constraints, economy-wide policy rules, sectoral demand, or other conditions relating aggregate allocations to idiosyncratic prices observed by the decision makers. $G(L)$ is a primitive of the model, and generates signals by

$$S_{N,i,t} = G(L)X_t \tag{11}$$

The right hand side of (11) includes no idiosyncratic terms, so $S_{N,i,t}$ is the same for all agents; it is determined only by macroeconomic aggregates X_t.

2.2.2 The Wold Representation

Before describing aggregation, it is necessary to characterize how white noise innovations $W_{i,t}$ are determined by the fundamental shocks $\varepsilon_{i,t}$.

The signal $A_{i,t}$ is equivalent to two polynomials: $S(L)\varepsilon_{i,t}$ is a lag operator polynomial of fundamental shocks, while *Wold representation* $A(L)W_{i,t}$ is a lag operator polynomial of white noise forecast errors. These white noise *Wold innovations* innovations can be written as

$$W_{i,t} = A(L)^{-1}S(L)\varepsilon_{i,t} \equiv W(L)\varepsilon_{i,t}$$

which has a variance matrix denoted by Σ_W.

Appendix C.4 describes how to compute the Wold representation from the signal autocovariance function. Let the $m_A \times m_A$ matrix Γ_j denote the jth autocovariance of the signal $A_{i,t}$. The fundamental shock $\varepsilon_{i,t}$ is a white noise process with unit variance, so the autocovariance Γ_j is given by

$$\Gamma_j = \sum_{k=0}^{\infty} S_k S'_{k-j} \tag{12}$$

2.2.3 Aggregation

Aggregate variables affect the endogenous signal, so I must characterize how shocks aggregate, and how aggregated shocks determine aggregate allocations.

The shock $\varepsilon_{i,t}$ contains both aggregate and idiosyncratic dimensions. Suppose there is a unit measure λ of agents i in the set $\mathcal{I}$. Assume the idiosyncratic dimensions are mean zero in the population. Then the average signal $A_t \equiv \int_{\mathcal{I}} A_{i,t} d\lambda(i)$ satisfies

$$A_t = \int_{\mathcal{I}} S(L)\varepsilon_{i,t} d\lambda(i)$$

because $S(L)\varepsilon_{i,t}$ is linear in the sequence of shocks. Similarly, the aggregate endogenous vector $X_t \equiv \int_{\mathcal{I}} X_{i,t} d\lambda(i)$ satisfies

$$X_t = \int_{\mathcal{I}} X(L) W_{i,t} d\lambda(i) = \int_{\mathcal{I}} X(L) W(L) \varepsilon_{i,t} d\lambda(i)$$
$$= X(L) W(L) \int_{\mathcal{I}} \varepsilon_{i,t} d\lambda(i) \quad (13)$$

Finally, let the projection matrix P_G denote the diagonal matrix with ones in dimensions corresponding to aggregate shocks and zeros elsewhere, so that

$$\int_{\mathcal{I}} \varepsilon_{i,t} d\lambda(i) = P_G \varepsilon_{i,t} \quad \forall i \in \mathcal{I} \quad (14)$$

2.2.4 The Fixed Point Equation

The lag operator polynomial $S_N(L)$ is determined by combining equations (11), (13), and (14):

$$S_N(L) = G(L) X(L) W(L) P_G \quad (15)$$

Adding $S_X(L)$ to equation (15) yields the signal process implied by equation (10). This defines the *Signal Operator* $\mathcal{B}(S(L))$:

$$\mathcal{B}(S(L)) \equiv S_X(L) + G(L) X(L) W(L) P_G \quad (16)$$

which can be written as a function of $S(L)$ because the polynomials $X(L)$ and $W(L)$ are both determined by $S(L)$. For cleanliness, I drop henceforth the (L) notation when writing the signal operator.

Combining equations (10), (15), and (16) provides the *equilibrium fixed point equation*:

$$S = \mathcal{B}(S) \quad (17)$$

The fixed point equation (17) states that the dynamics of the signal process must be consistent with the dynamics it implies for the endogenous variables.[10] If the signal S satisfies the fixed point equation (17), then it is said to be an *equilibrium signal process*.

The next sections and main results of the paper are focused on understanding the operator $\mathcal{B}$, how to compute it, and what can be said about its fixed points.

[10]Why are there no higher order expectations in the fixed point equation? In many models (e.g. the beauty contests in Keynes (1936), or more recently in papers such as Morris and Shin (2002), Woodford (2003), Allen, Morris, and Shin (2006), Makarov and Rytchkov (2012), or Nimark (2017)) agents must forecast the forecasts of others, which themselves depend on forecasts of forecasts, and so on, leading to a hierarchy of higher order expectations. Explicitly finding this hierarchy is challenging, but fortunately not necessary in general to solve for rational expectations equilibria. Instead, it sufficient to require that all agents make their best possible forecasts given their information sets. The fixed point equation does exactly this. This insight has been understood since at least Townsend (1983), and Huo and Pedroni (2020) solve for expectations in a general class of beauty contest models using this approach.

2.3 A Simple Example: Asset Pricing with Confounding Dynamics

Throughout the theoretical sections that follow, it is useful to see the theorems applied to a common example. So this section introduces a model of asset pricing with dispersed information and shows how to represent it in the form used in this paper. Then later sections (3.2, 3.4, and 3.5) will discuss its theoretical properties further after new theorems are introduced. Section 4.2 modifies the model to include additional idiosyncratic noise.

2.3.1 Model Assumptions

This asset pricing model features "confounding dynamics" (Rondina and Walker, 2021). There are as many shocks as signals, so it is possible for endogenous signals to reveal full information. But there are also other equilibria where endogenous behavior confounds inference of the underlying shocks. The regularity condition and signal-stability uniqueness theorem help distinguish between these equilibria. Specifically: full information is the *only* signal-stable equilibrium.

Agents forecast the value of an asset in the next period, discounted by a factor $\beta \in (0,1)$. The fundamental value of the asset is x_t, which is determined stochastically:

$$x_t = F(L)u_t$$

with standard normal shocks $u_t \sim N(0,1)$ and rational $F(L)$. Agents' choice variable is their discounted forecast:

$$p_{i,t} = \beta E_{i,t}[x_{t+1}]$$

Agents do not observe the current value x_t exactly, but they do see a noisy signal $z_{i,t}$ with standard normal idiosyncratic error $y_{i,t}$:

$$z_{i,t} = x_t + y_{i,t}$$

and they also observe the average price p_t:

$$p_t = \int_{i \in \mathcal{I}} p_{i,t} di$$

2.3.2 Representation in the General Framework

How does this model map to the general form? The signal vector is $A_{i,t} = \begin{pmatrix} z_{i,t} \\ p_t \end{pmatrix}$. The equilibrium condition is conveniently rewritten as $p_{i,t} = \beta E_{i,t}[z_{i,t+1}]$ because forecasting x_{t+1} is equivalent to forecasting $z_{i,t+1}$. In this case, mapping to the form in equation (1) implies $B_{X0} = -1$, $B_{X1} = 0$, $B_{A0} = \begin{pmatrix} 0 & 0 \end{pmatrix}$ and $B_{A1} = \begin{pmatrix} \beta & 0 \end{pmatrix}$

The exogenous signal process for this model is $S_X(L)\varepsilon_{i,t} = \begin{pmatrix} F(L) & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} u_t \\ y_{i,t} \end{pmatrix}$, while the endogenous signal process is

$$S_N(L)\varepsilon_{i,t} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} p_t = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left[L^{-1} \begin{pmatrix} \beta & 0 \end{pmatrix} A(L)\right]_+ W(L) P_G \tag{18}$$

with $P_G = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ to identify only the aggregate shock u_t from the vector $\varepsilon_{i,t}$. This model is simple because there is no dynamic interaction between endogenous variables; the feedback operators are $\Theta\Xi = I$ and $G = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

2.3.3 Equilibria in the Confounding Dynamics Model

One benefit of this particular example is that the solutions are known without resorting to numerical methods.

Under full information, the equilibrium is $p_{i,t} = \beta f_t$, where f_t denotes the full information forecast of x_{t+1}:

$$f_t \equiv [L^{-1}F(L)]u_t$$

This is always an equilibrium of the incomplete information model as well: if agents observe $p_t = \beta f_t$, then they form expectations by $E_{i,t}[x_{t+1}] = f_t$.

But if $F(L)$ is noninvertible, then there exists a second equilibrium (or more) featuring "confounding dynamics". Denote the Wold representation of $F(L)x_t$ by

$$F(L)x_t = A^F(L)w_t^F$$

with causal and invertible $A^F(L)$ and white noise w_t^F. Furthermore, assume that $F(L)$ is such that the forecast polynomial is $[L^{-1}A^F(L)]_+$ is invertible (e.g. $F(L)$ could be MA(1) or ARMA(1,1)).

Proposition 1 *If $F(L)u_t$ is noninvertible with forecast errors w_t^F then $p_t^{CD} \equiv \beta[L^{-1}A^F(L)]_+ w_t^F$ is an equilibrium price process of the confounding dynamics model.*

Proof: Appendix A.11.1

Full information and confounding dynamics are both equilibria of this model. But as Section 3.5 shows, both equilibria are not stable.

3 Properties of the Signal Operator

In this section, I introduce the Signal Operator Iteration algorithm, a fixed point of which solves the general model of Section 2. I define the *Information Feedback Regularity* condition, and show that it is necessary for signal-stable equilibria, which are locally unique and have convergent sequences of approximations. Finally, I give the global uniqueness theorem for signal-stable equilibria.

But first, I introduce notation and describe how to represent signals as Toeplitz operators.

3.1 Block Toeplitz Operator Representation of Lag Operator Polynomials

In order to characterize fixed points, it is useful to treat signal processes and other lag operator polynomials as bounded linear operators on a Hilbert space. This is helpful because operator properties are valuable for proving and understanding properties of equilibrium. For example, the Information Feedback Regularity condition introduced in Section 3.2 is defined in terms of the operator norm, and proving the uniqueness theorems repeatedly uses operator properties. This representation is also helpful because it maps directly to a computational strategy, as Appendix C makes clear. Working with operators brings some additional notation, but the mathematics are familiar: the bounded linear operators are simply infinite-dimensional matrices.

Specifically, an arbitrary $n \times m$ square summable lag operator polynomial $Y(L) = \sum_{j=-\infty}^{\infty} Y_j L^j$ is a bounded linear operator on an infinite sequence of shocks or another time series. $Y(L)$ has a representation as a block *Toeplitz* operator, which is the infinite analog to a block Toeplitz matrix. For notation, let Y denote the Toeplitz operator of the polynomial $Y(L)$, which maps $\ell^2 \to \ell^2$.[11] The operator Y has $n \times m$ blocks, so Y maps $m \times 1$ shocks to $n \times 1$ signals. For the arbitrary operator Y, the matrix form is:

$$\begin{pmatrix} Y_0 & Y_{-1} & Y_{-2} & Y_{-3} & \ldots \\ Y_1 & Y_0 & Y_{-1} & Y_{-2} & \ldots \\ Y_2 & Y_1 & Y_0 & Y_{-1} & \ldots \\ Y_3 & Y_2 & Y_1 & Y_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{19}$$

The product of two operators is the product of the infinite Toeplitz matrices, the inverse Y^{-1} of the operator is the inverse of the infinite Toeplitz matrix, and so forth.[12] The lag operator L is the Toeplitz operator with identity matrices along the first block diagonal below the main block diagonal. When $Y(L)$ is *causal*, so that it has $Y_j = 0$ for all $j < 0$, the operator Y is *lower block triangular.*

When $Y(L)$ is a constant matrix so that $Y_j = 0$ for all $j \neq 0$, then the operator Y is block diagonal with Y_0 along the main block diagonal. To ease notation, I let the matrix Y_0 also denote its corresponding block diagonal operator, so that I do not

[11] Appendix E elaborates on how to represent time series in this space.

[12] For this and other useful properties of operators on ℓ^2, see Conway (2007), or Frazho and Bhosri (2010) for Toeplitz operators in particular.

have to define a new operator for every matrix that is added to or multiplied by a lag operator polynomial.

The signal operator $\mathcal{B}(S)$ is a nonlinear operator, mapping $\mathcal{S}_{m_A,m_\varepsilon} \to \mathcal{S}_{m_A,m_\varepsilon}$.[13] $\mathcal{S}_{m_A,m_\varepsilon}$ denotes the Banach space of *causal* block $m_A \times m_\varepsilon$ Toepliz operators that map m_ε-dimensional random shocks to m_A-dimensional signals. $\mathcal{S}_{m_A,m_\varepsilon}$ is a Banach space, and the distance metric on this space is the norm $\|\cdot\|_S$.[14]

3.2 The Regularity Condition

Will the Signal Operator Iteration algorithm be well behaved? Can a fixed point be stable?

These questions are answered by evaluating whether the model satisfies a regularity condition: *Information Feedback Regularity* (IFR). The condition characterizes the potential size of the information feedback in the model. If the feedback is small from signals to decisions to signals, then it is possible for stable equilibria to exist such that Signal Operator Iteration will converge given a good initial guess. The condition is given by:

Condition 1 *A model satisfies Information Feedback Regularity if*

$$\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| < 1$$

$\|\cdot\|$ here denotes the operator norm, which measures by how much the operator can increase the variance of any signal process. Accordingly, Condition 1 says that this operator must decrease the variance of a signal. Because the operator maps ℓ^2 to ℓ^2, the operator norm is the largest singular value, which is analogous to the matrix norm in finite dimensions, and can be easily computed.

The non-causal operator $G\Theta\Xi(B_{A1}L^{-1} + B_{A0})$ depends entirely on the primitives of the model in question. Therefore IFR can be evaluated without solving the model. The norm of this operator represents how much $\mathcal{B}(S)$ can be changed by perturbing the signal process S in a way that is spanned by a forecast error process W.[15] To see why, write out $\mathcal{B}(S)$ using Theorem 1 (and omitting the (L) notation):

$$\mathcal{B}(S) = S_X + G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ WP_G \tag{20}$$

Two components make up the feedback mechanism. The first component G determines how aggregate actions affect individuals' signals; when entries in G are large,

[13] Lemma 11 proves this self-map.

[14] Appendix A.2 defines this space and norm formally.

[15] The annihilator $[\cdot]_+$ does not appear in the IFR definition because the non-causal components of $G\Theta\Xi(B_{A1}L^{-1}+B_{A0})$ are relevant for perturbations linear in W but with very long lags; Theorem 3 shows that the entire non-causal operator gives the relevant necessary condition. Perturbations orthogonal to W are more complicated: instead of a single expression, the relevant norm depends on the signal S around which the perturbation is made. This is why calculating the general Fréchet derivative of $\mathcal{B}$ (Theorem 8) is much more challenging.

small changes in actions have large effects on the information process. The second component $\Theta\Xi(B_{A1}L^{-1} + B_{A0})$ determines how information maps to actions via the policy function (Theorem 1); when this operator is large, small changes in the information agents observe have large effects on their actions. Condition 1 may be violated if either of these terms is too large. For example, if the feedback from information to actions to information is via an inelastic channel (such as capital) the feedback may be small so that the condition is satisfied. However, if the feedback is via a very elastic channel so that the regularity condition is not satisfied, several problems occur: an equilibrium will not be signal-stable (Theorem 3), and may not be solvable (Corollary 1).

Information Feedback Regularity in the Confounding Dynamics Example

The regularity condition is easy to check in the asset pricing example (Section 2.3):

Proposition 2 *Information Feedback Regularity is satisfied in the confounding dynamics model if* $\beta \in (0, 1)$.

Proof: Appendix A.11.1

The discount factor β controls the information feedback in the confounding dynamics model. When β is large, then small changes in average expectations have large effects on the endogenous signal. But when $\beta < 1$, IFR holds, dampening the feedback from expectations to signals to expectations.

3.3 Approximate Fixed Points

Fixed points of the true model (17) may be infinite-dimensional and uncomputable.[16] It is helpful to first understand computable *approximate* fixed points of a finite approximation of the model $\mathcal{B}_\tau$. This is useful for characterizing numerical solution methods, but this approximation is also used in the proofs of several theorems regarding the true model.

The finite approximation $\mathcal{B}_\tau$ is defined:

$$\mathcal{B}_\tau(S) \equiv \mathcal{B}(S)P_\tau \tag{21}$$

The projection operator P_τ truncates a signal process after lag τ, which I refer to as the "order" of the approximation. This is a standard approach to approximating infinite-dimensional Toeplitz operators known as the "finite section method" (Böttcher and Silbermann, 2012). Fortunately, even though the true equilibrium may be of infinite order, it can be approximated arbitrarily well with the finite section method; Theorem 2 formalizes this property.

[16] See Sargent (1991), Makarov and Rytchkov (2012), or Huo and Takayama (2023) among others.

The solution algorithm *Signal Operator Iteration* repeatedly applies $\mathcal{B}_\tau$ to find a fixed point, so that guesses of the signal S^n and S^{n+1} are related by

$$\begin{aligned} S^{n+1} &= \mathcal{B}_\tau(S^n) \qquad (22) \\ &= \left(S_X + GX^n W^n P_G\right) P_\tau \end{aligned}$$

Appendix C describes this algorithm in detail and explains how to compute it numerically.

A fixed point $\hat{S}_\tau$ of $\mathcal{B}_\tau$ is called an *approximate fixed point* (or colloquially, an *approximate solution of order* τ) of the macroeconomic model, satisfying:

$$\hat{S}_\tau = \mathcal{B}_\tau(\hat{S}_\tau) \tag{23}$$

Finite-dimensional signals may be true solutions (for example, the beauty contest models in Section 5.1) but in many cases will only be solutions to the finite approximation $\mathcal{B}_\tau$ of the true model $\mathcal{B}$. A finite truncation is sufficient for all practical purposes because an infinite dimensional object is never computable. Still, it may be valuable to know whether the model has a solution without truncation, and whether the finite solution is a good approximation to the infinite case. Theorem 2 affirms this to be true.

The fixed point $\mathcal{B}(\hat{S}) = \hat{S}$ in equation (17) is the equilibrium fixed point of the true Signal Operator defined in equation (16). Theorem 2 proves that if finite-dimensional solutions converge, the limiting signal process is the fixed point of the infinite-order signal operator. This is because the finite-order signal operator can approximate the uncomputable infinite-order signal operator arbitrarily well (Lemma 3).

Theorem 2 (Limits of Approximate Fixed Points) *If $\hat{S}_\tau$ is a sequence of fixed points satisfying $\mathcal{B}_\tau(\hat{S}_\tau) = \hat{S}_\tau$ and $\lim_{\tau\to\infty} \hat{S}_\tau = \hat{S}$, then $\hat{S}$ is a fixed point of the infinite-order signal operator, i.e. $\mathcal{B}(\hat{S}) = \hat{S}$*

Proof: Appendix A.4

What is the practical implication of Theorem 2? If the model has an approximate solution for large τ, and further increases to τ make little difference to the numerical solution, then the infinite-order solution must exist, and can be approximated arbitrarily well.[17]

Conversely, if an infinite-order true solution exists, will there necessarily be a sequence of finite approximates that converges to it? Yes, if the true solution is *signal-stable*. The next section defines signal-stability and explores its properties.

[17] The theorem also demonstrates an advantage of this signal truncation after τ lags, compared to revealing the entire set of underlying shocks after τ periods. A limit of solutions under this alternative approach may not be the solution to the true model.

3.4 Equilibrium Stability and Uniqueness

In this section I study signal-stable equilibria and local uniqueness. I show *Information Feedback Regularity* is necessary for signal-stable equilibria to exist.

Why do we care about stability? Signal-stable equilibria are interesting because they are robust to small perturbations, and can be found numerically. They are also informed by Information Feedback Regularity: they cannot exist without IFR. And most importantly, signal-stable equilibria are *globally unique*.

3.4.1 Signal-Stability

Any equilibrium signal process $\hat{S}$ is a fixed point satisfying $\hat{S} = \mathcal{B}(\hat{S})$. In general, the set of possible equilibria is difficult to characterize because $\mathcal{B}$ is nonlinear. However, it is possible to characterize a refined set of equilibria with an important property: *signal-stability*.

Definition 1 *An equilibrium fixed point signal satisfying $\hat{S} = \mathcal{B}(\hat{S})$ is called* signal-stable *if there exists some neighborhood of $\hat{S}$ such that for any S^{Δ} in the neighborhood, $\|\mathcal{B}(S^{\Delta}) - \mathcal{B}(\hat{S})\|_S < \|S^{\Delta} - \hat{S}\|_S$. Otherwise, $\hat{S}$ is called* signal-unstable.

What characterizes a signal-stable equilibrium? If you perturb the signal process, the change in forecasts will not be so large that the implied endogenous signal changes by more than the perturbation. Signal-unstable equilibria are typically cases where small perturbations are explosive, but also the edge cases, where small perturbations of input signals imply an equal perturbation of output signals. This type of stability is known in other settings as "contractive stability".[18]

The term "signal-stability" emphasizes that this concept refers to stability with respect to perturbations to the *signal process*, as opposed to stability regarding only policy functions or economic outcomes. Because the signal vector may include exogenous signals, a signal-stable equilibrium must be robust to perturbations in these dimensions as well. In either case, the perturbation is not to the exogenous stochastic process driving the economy (S_X is always unchanged). Signal perturbations can be thought of as random noise in the signals observed by agents, which may be driven by either extrinsic sunspots or fundamental shocks.

Why are signal-stable equilibria interesting? They are robust to small changes in the information process. In models with endogenous signals, the endogenous component depends on the equilibrium signal itself. If an equilibrium is signal-stable, this self-referential feedback is well behaved. If an equilibrium is signal-unstable, this feedback is explosive, so that small perturbations in the signal process can produce ever larger perturbations in the endogenous component, diverging away from the equilibrium. As a practical matter, signal-unstable equilibria cannot necessarily be found

[18] This is because the operator $\mathcal{B}$ is a contraction on a neighborhood around a signal-stable fixed point. Signal-stability is stronger than "exponential stability" which only requires that $\mathcal{B}^k$ is contraction, for some possibly large k. Appendix B discusses this latter form of stability.

by iterative methods. As a conceptual matter, the explosive sensitivity could make signal-unstable equilibria unlikely to be observed in the real world, where modeling error or other perturbations appear.[19] Still, signal-unstable equilibria are valid model solutions and I can neither rule them out nor characterize them in general.

Signal-stability is a more powerful property than *solvability*. If a model is solvable by iterative methods it is not necessarily signal-stable and signal-stability guarantees local uniqueness (Theorem 4) while solvability does not. Applying Signal Operator Iteration or another algorithm to solve a model does not guarantee that the solution is well behaved, even if the algorithm converges: it is possible to converge in some region around a fixed point, but diverge in another. For example, a saddle point features this property, where the saddle path will converge to the point, but every other point in any neighborhood around the solution will diverge. Or it is possible to have a path that converges to an continuous connected region of valid fixed points, such that iterative algorithms will converge to points on the boundary, which are surrounded by other valid solutions. The possibility of such solutions is compounded by the infinite dimensional nature of the space, where any particular dimension might be one in which perturbations lead to divergence or alternative fixed points. Signal-stability rules out such possibilities.

The method to check if a fixed point $\hat{S}$ is signal-stable is by evaluating the norm of the Fréchet derivative $D_{\mathcal{B}}(\hat{S})$ at that point

Property 1 *$\hat{S}$ is a signal-stable equilibrium if and only if $\|D_{\mathcal{B}}(\hat{S})\| < 1$.*

This property follows directly from the definition of signal-stability: the Fréchet derivative is the operator-valued derivative of $\mathcal{B}$, so the norm $\|D_{\mathcal{B}}(\hat{S})\|$ is the size of the greatest marginal deviation of $\mathcal{B}$ around $\hat{S}$. Evaluating stability by calculating the Fréchet derivative is analogous to evaluating the stability of the fixed point x of a scalar-valued function f by calculating $f'(x)$. Theorem 8 gives the exact expression for $D_{\mathcal{B}}(\hat{S})$.

3.4.2 Stability and the Regularity Condition

Information Feedback Regularity determines if a model's feedback from information to actions to information can be explosive. So it is intuitive that signal-stability should depend in some way on the regularity condition. Theorem 3 states that regularity is a necessary condition for signal-stability when the model has a typical feature: if an equilibrium signal process must include an aggregate signal. This is common in

[19] When considering signal-stable or unstable equilibria of dynamic systems, economists may be reminded of the neoclassical growth model, where the equilibrium is a saddle path. Economists must be careful not to conclude that signal-unstable equilibria are robust in other models: transversality and resource constraints rule out any explosive paths in the neighborhood of the neoclassical growth equilibrium, but this is not the case in general. In this paper, there are no general assumptions to rule out a perturbation that could put the endogenous signal process on an explosive path away from an unstable equilibrium towards a stable one.

macroeconomic models with information frictions; agents observe something about the aggregate economy even though they do not observe everything about it.[20]

Theorem 3 (The Necessary Condition) *If all fixed points of a model contain aggregate signals such that for any fixed point signal vector $\hat{S}$ there is an entry $\hat{S}_i$ satisfying $\hat{S}_i P_G = \hat{S}_i$ then Information Feedback Regularity is a necessary condition for signal-stable fixed points to exist.*

Proof: Appendix A.6

Does a signal-stable equilibrium exist? It is difficult to tell *a priori* because the Fréchet derivative is generally unbounded on $\mathcal{S}_{m_A,m_\varepsilon}$, due to the signal inverses that appear in Theorem 8. Indeed, this is why uniqueness cannot be guaranteed in some otherwise well-behaved full information models once information endogeneity is considered (Adams, 2022). As the next sections demonstrate, signal-stable equilibria have many desirable characteristics, and Information Feedback Regularity is a necessary condition. This property can be helpful: a practitioner can easily check IFR without solving the model see if they have any hope of finding a signal-stable equilibrium. Section 5 gives several examples of this application.

3.4.3 Useful Properties of Stable Fixed Points

One valuable property of signal-stable fixed points is that the operator $\mathcal{B}$ is a contraction near any fixed point. This implies that a fixed point is locally unique, which is a weaker property than signal-stability: all signal-stable equilibria must be locally unique, but the converse is not necessarily true. Theorem 4 formalizes this property.

Theorem 4 (Local Uniqueness) *If $\hat{S} \neq 0$ is a signal-stable fixed point of $\mathcal{B}$, then $\mathcal{B}$ is a contraction on a neighborhood around $\hat{S}$, and $\hat{S}$ is a locally unique fixed point.*

Proof: Appendix A.6

Another valuable property is that repeatedly applying the signal operator $\mathcal{B}$ to any guess S_0 that is sufficiently close to a signal-stable fixed point will necessarily converge to the fixed point. Corollary 1 states this formally.

Corollary 1 *If $\hat{S}$ is a signal-stable fixed point, then there exists a neighborhood around it $b(\hat{S})$ such that for any point $S_0 \in b(\hat{S})$*

$$\lim_{k\to\infty} \mathcal{B}^k S_0 = \hat{S}$$

[20]In most applications this condition is straightforward to check. If an exogenous signal with idiosyncratic elements is added to all non-zero endogenous signals, then the condition will fail. Otherwise, one can check the condition by determining if zero-valued endogenous signals imply non-zero-valued endogenous signals, (i.e. if $\mathcal{B}(S_X) \neq S_X$) thus ruling out fixed points without aggregate signals. This condition is needed to prove necessity in Theorem 3 because it implies that there is a dimension of the fixed point signal that the aggregating operator P_G does not shrink. Thus, there are dimensions for which the norm $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|$ alone determines if vector lengths must decrease or not.

Proof: Appendix A.6

Stability is also helpful for understanding uncomputable infinite-dimensional equilibria. Theorem 5 says that if such an equilibrium exists and is signal-stable, then it is the limit of a sequence of approximate signal-stable equilibria.

Theorem 5 (Computability) *If $\hat{S} = \mathcal{B}(\hat{S})$ is a signal-stable fixed point, then there exists a sequence of signal-stable approximate fixed points $\hat{S}_\tau = \mathcal{B}_\tau(\hat{S}_\tau)$ such that*

$$\lim_{\tau \to \infty} \hat{S}_\tau = \hat{S}$$

Proof: Appendix A.6

This result is a companion to Theorem 2, which had a practical application: if a practitioner found a converging sequence of approximate solutions, they could be confident that the limit was the true solution. In contrast, Theorem 5 implies that if there is a true *signal-stable* solution, a converging sequence of approximations exists, and the elements of that sequence will also be signal-stable.

Stability in the Confounding Dynamics Example

Multiple equilibria are possible in this example (Section 2.3). Are there any properties that might lead a practitioner to prefer one? Yes, Proposition 3 states that the full information equilibrium is always *signal-stable* if Information Feedback Regularity is satisfied.

Proposition 3 *The full information equilibrium of the confounding dynamics model $p_t = \beta f_t$ is signal-stable if $\beta \in (0, 1)$.*

Proof: Appendix A.11.1

The full information equilibrium is necessarily signal-stable.[21]

3.5 Global Uniqueness of Stable Equilibria

Is an equilibrium globally unique? In general, this is a difficult question to answer, because the mapping from information to actions to information is highly nonlinear when a model features information frictions. However, it is possible to prove uniqueness within a useful class: signal-stable equilibria. Theorem 6 states this result.

Theorem 6 *If Information Feedback Regularity holds, then there exists at most one signal-stable fixed point of $\mathcal{B}$.*

[21] This is true for more general models as well: the set of invertible operators is open, so small deviations from signals that reveal full information will still reveal full information. And if IFR holds, the full information equilibrium will be signal-stable.

Proof: Appendix A.7

Theorem 6 applies to the true signal operator $\mathcal{B}$, as well as any finitely approximated $\mathcal{B}_\tau$.

Again, Information Feedback Regularity is the crucial property: when it holds, there exists at most one signal-stable equilibrium. General proofs of uniqueness are challenging because of the nonlinearity of the signal operator $\mathcal{B}$ and non-compactness of the signal space. If practitioners care about signal-stable equilibria *a priori*, this theorem allows practitioners to guarantee a unique solution within this class by satisfying Information Feedback Regularity.[22]

The proof is topological. It begins by defining a space $\mathcal{Y}_\tau$ in which $I - \mathcal{B}_\tau$ is a local homeomorphism and which must contain all signal-stable approximate fixed points. Then the proof demonstrates that this space is path connected, and the operator $I - \mathcal{B}_\tau$ is a global homeomorphism on this space, ensuring a unique fixed point. Finally, Theorems 4 and 5 extend the result to all signal-stable fixed points, even those that are infinite-dimensional.

Signal-Stability and Uniqueness in the Confounding Dynamics Example

Proposition 3 said that the full information equilibrium in the simple example (Section 2.3) was signal-stable. But there are multiple equilibria. Are the other equilibria signal-unstable? Yes.

This is a direct consequence of Theorem 6: the full information equilibrium of the confounding dynamics model $p_t = \beta f_t$ is the *unique* signal-stable equilibrium. Appendix D.2 demonstrates constructively in an example that even when IFR holds, the confounding dynamics equilibrium is indeed signal-*unstable*. Specifically, there is a small perturbation to the confounding dynamics equilibrium such that Signal Operator Iteration will eventually diverge and approach the full information solution.

There is a lesson from this proposition for practitioners. In models with multiple equilibria and confounding dynamics, signal-stability can inform equilibrium selection if Information Feedback Regularity is satisfied.

4 A Sufficient Condition for Signal-Stability

Information Feedback Regularity is a necessary condition for global uniqueness. This section provides a sufficient condition: *Sufficient Idiosyncrasy.* The condition ensures that all fixed points are signal-stable, of which Theorem 6 says there can be at most one.

The central insight is that for a model with enough idiosyncratic shocks, as the idiosyncratic variance becomes large, it dampens the information feedback. In the

[22]Moreover, if all fixed points of a model must contain aggregate signals, Theorem 3 applies and the IFR requirement in Theorem 6 is redundant: without IFR, no signal-stable fixed points can exist.

limit, as the relative variance of aggregate to idiosyncratic shocks goes to zero, the model approaches an exogenous information model, which has a unique solution so long as the usual Blanchard-Kahn condition is satisfied (Theorem 1).[23] Theorem 7 in this section shows that there is a uniqueness region where idiosyncratic noise is sufficiently large but finite.

The condition is strong and may not hold in many models. In particular, it requires that *all* signals are subject to idiosyncratic noise. But if the condition holds, it guarantees that a model has a unique equilibrium.

4.1 The Sufficient Idiosyncrasy Condition

Consider $S_{X,0}(I - P_G)$, the contemporaneous component of the exogenous idiosyncratic process. In this expression, S_X is the exogenous process, $S_{X,0}$ is the matrix corresponding to contemporaneous shocks, and the matrix $I - P_G$ isolates the idiosyncratic dimensions. The signal variance due to contemporaneous idiosyncratic shocks is therefore given by the matrix

$$\Sigma_I \equiv S_{X,0}(I - P_G)S^*_{X,0} \tag{24}$$

If this variance matrix is invertible, let $\mathrm{r}(\Sigma_I^{-1})$ denote the spectral radius of the inverse, i.e. the smallest eigenvalue of Σ_I. If the variance is not invertible, then as a convention write $\mathrm{r}(\Sigma_I^{-1}) = \infty$ even though strictly speaking, Σ_I^{-1} does not exist.

Some additional notation is worth reviewing for the sufficient condition. First, the norm of the information feedback is $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|$. When this norm is less than one, IFR is satisfied. Secondly, the "forecast radius" R_N is described in Lemma 2 as a bound on the endogenous signal S_N. The radius is given by

$$R_N = \frac{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|}$$

where

$$\vartheta_I = \|G\Theta\|\| \left[\Xi(B_{A1}L^{-1} + B_{A0})S_X(I - P_G)\right]_+ \|_S$$

is similar to the transformation generating the endogenous signal S_N, except where the idiosyncratic component $S_X(I - P_G)$ is forecasted instead of the aggregate component SP_G in equation (15).

Condition 2 *A model satisfies Sufficient Idiosyncrasy if*

$$r(\Sigma_I^{-1}) < \frac{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2}{4R_N^2}$$

[23]Levine, Pearlman, Wright, and Yang (2024) use a related property to study the behavior of endogenous signal models in the limit as idiosyncratic noise becomes infinitely large.

The spectral radius $\mathrm{r}(\Sigma_I^{-1})$ is positive, so the Sufficient Idiosyncrasy Condition (SIC) implies that Information Feedback Regularity must hold: $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| < 1$.

SIC also implies that there are at least as many idiosyncratic shocks as there are signals, because it is necessary for Σ_I to be invertible. In what types of settings will this hold? Models where the endogenous signals are aggregates – such as asset prices in the confounding dynamics example – will necessarily fail this condition. However, if agents are forecasting exclusively using idiosyncratic variables then some parameterization will satisfy SIC.[24] In the next section I demonstrate this by modifying the confounding dynamics model so that both signals are distorted by idiosyncratic noise.

But first, why is the SIC useful? It ensures equilibrium uniqueness:

Theorem 7 *If the Sufficient Idiosyncrasy Condition holds, then if a model has a fixed point, it is the globally unique fixed point.*

Proof: Appendix A.8

Theorem 7 says that the SIC is a sufficient condition for global uniqueness.This is because SIC guarantees that any fixed point is signal-stable. It also guarantees that IFR holds, so Theorem 6 implies that there can be at most one signal-stable fixed point.

4.2 The Sufficient Idiosyncrasy Condition in the Confounding Dynamics Model

To demonstrate how the Sufficient Idiosyncrasy Condition (SIC) guarantees a globally unique fixed point, this section modifies the confounding dynamics model discussed thus far.

SIC fails in the original model introduced in Section 2.3. This is because there are two signals but only one idiosyncratic shock, so the variance matrix Σ_I is uninvertible.

Therefore, I modify the model by introducing idiosyncratic noise $\epsilon^v_{i,t} \sim N(0,1)$ to agents' observation of the price. The two signals that agents observe are now

$$z_{i,t} = x_t + y_{i,t} \qquad\qquad s_{i,t} = p_t + \tau_v^{-\frac{1}{2}}\epsilon^v_{i,t}$$

where $\epsilon^v_{i,t}$ is entirely idiosyncratic. The standard deviation of the noise is $\tau_v^{-\frac{1}{2}}$, which is convenient to express in terms of the noise precision τ_v. In order to have a simple analytical solution in this example, I assume the fundamental value of the asset is given by

$$x_t = (1 + \alpha L)u_t$$

[24]Examples of models with idiosyncratic signals alone include forecasting from firm-level productivity in Venkateswaran (2014), house prices in Chahrour and Gaballo (2020), household income in Adams and Rojas (2024), or local interest rates in Angeletos and Lian (2020), among many others.

where $\alpha > 1$ and $u_t \sim N(0,1)$, which matches the example in Appendix D. This is an uninvertible MA(1), whose Wold representation is

$$x_t = (1+\theta L)w_t^x$$

where $\theta = \alpha^{-1}$ and the white noise process is $w_t^x = \alpha\frac{\theta+L}{1+\theta L}u_t$.

Let the noise term $y_{i,t}$ have a similar form:

$$y_{i,t} = (1+\theta L)\tau_y^{-\frac{1}{2}}\epsilon_{i,t}^y$$

where again $\epsilon_{i,t}^y$ is standard normal, and the noise standard deviation is written in terms of the precision parameter τ_y.

Because the price signal $s_{i,t}$ now has idiosyncratic noise, full information is no longer an equilibrium of this model. But Proposition 4 states that there is at least one equilibrium whose aggregate price is proportionate to the confounding dynamics equilibrium of the original model.

Proposition 4 *The modified confounding dynamics model features an equilibrium where the aggregate price $\bar{p}_t$ is given by*

$$\bar{p}_t = bw_t^x$$

for some b, and b is unique if and only if

$$27\beta^2\tau_v\tau_y^2 > 4(\beta\tau_v - \tau_y - 1)^3 \tag{25}$$

Proof: Appendix A.9

Proposition 4 makes it clear that the precision of the idiosyncratic noise is relevant for global uniqueness. The inequality (25) is satisfied only if τ_v is sufficiently large: when the precision of the idiosyncratic signal is high, then the model has multiple equilibria.

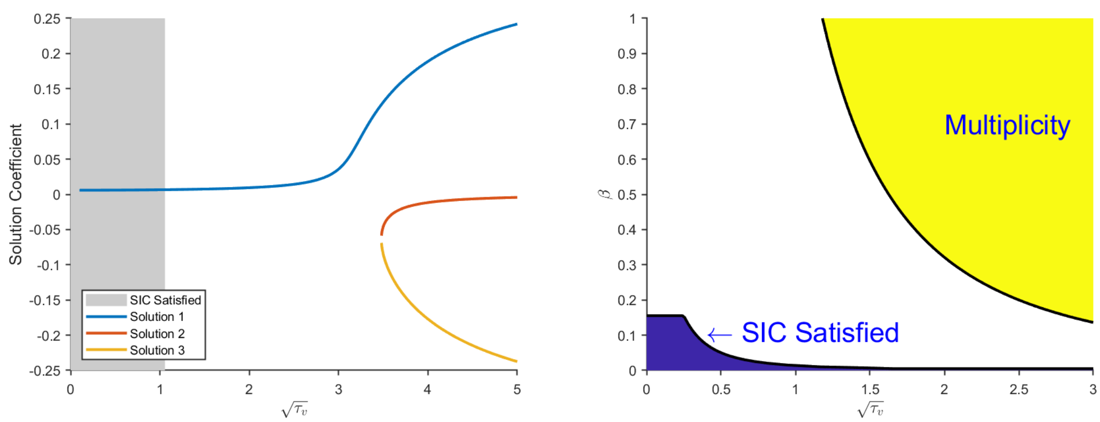
This is the force that the Sufficient Idiosyncrasy Condition constrains in the general class of models. The SIC requires that the idiosyncratic precision is sufficiently small to guarantee that all fixed points are stable. Proposition 5 demonstrates, giving the SIC for this modified confounding dynamics model. It shows that if the precisions τ_y and τ_v are sufficiently small, a unique equilibrium is guaranteed.

Proposition 5 *In the modified confounding dynamics model, the Sufficient Idiosyncrasy Condition is satisfied if*

$$\max\{\tau_y, \tau_v\} < \frac{(1+\beta)(1-\beta)^3}{4\beta^2\left(\sqrt{1+\alpha^2}+\theta\tau_y^{-1/2}\right)^2}$$

Proof: Appendix A.9

Figure 1 illustrates the relationship between idiosyncratic noise and multiplicity. The model parameters are $\alpha = 2$, $\sqrt{\tau_y} = .2$ (i.e. the standard deviation of $\epsilon^y_{i,t}$ is 5 times as large as the standard deviation of the fundamental shock u_t) and $\beta = 0.1$ so that the information feedback is limited. Panel (a) examines how the model solution depends on the root precision $\sqrt{\tau}_v$ (the inverse of the standard deviation of the idiosyncratic shock $\epsilon^v_{i,t}$). The y-axis measures the coefficient b in the solution $\bar{p} = bw^x_t$ described in Proposition 4. When the precision is low, Theorem 7 guarantees that there is a unique solution: the gray region indicates when the SIC holds. Conversely, when the precision is sufficiently large, the model has multiple equilibria.



(a) Multiplicity and SIC: Solutions

(b) Multiplicity and SIC: Regions

Figure 1: Equilibrium Uniqueness in the Confounding Dynamics Model with Idiosyncratic Noise

The left panel plots the solutions described by Proposition 4 for the modified confounding dynamics model with idiosyncratic noise. The model parameters are $\alpha = 2$, $\sqrt{\tau_y} = .2$, and $\beta = 0.1$; the figure shows how the solutions vary with the signal precision τ_v. The right panel also varies β on the y-axis. The yellow region has multiple equilibria; in the blue region the SIC is satisfied, so Theorem 7 ensures that the solution is unique.

There is a fundamental trade-off between idiosyncratic noise and information feedback. In general, when the information feedback $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|$ is large, then more idiosyncratic noise is needed to make $r(\Sigma_I^{-1})$ small enough for the SIC to hold. Panel (b) of Figure 1 illustrates this trade-off in the model. The calibration is unchanged from Panel (a), except now the information feedback parameter β is allowed to vary, and the figure plots the combinations of $\sqrt{\tau_v}$ and β that imply multiplicity or SIC. When the information feedback β is small, then SIC is easily satisfied, and there is only multiplicity for extremely high precisions. In the limit as $\beta \to 0$, there is no multiplicity at all. However, when β is larger, then the noisy signals must be less precise in order for the model to have a unique equilibrium. Eventually, a large enough information feedback implies that SIC cannot hold for any τ_v. Theorem 7 implies that the two regions in the figure can never overlap: if SIC is satisfied, the model must have a unique equilibrium.

5 Examples Applying Information Feedback Regularity

This section applies lessons from Section 3 to a number of additional simple examples motivated by the literature.[25] These examples demonstrate how to map a variety of different models into this paper's general form, how to determine if Information Feedback Regularity holds, and how to draw useful conclusions based on the answer.

5.1 Beauty Contests with Endogenous Information

This section studies a beauty contest model, modifying a structure resembling Angeletos and La'O (2010) with endogenous signals. The exercise demonstrates how to represent a static nowcasting problem in the general macroeconomic form of Section 2, and how to apply the theoretical results from Section 3 in a model with analytical solutions. Specifically, this model demonstrates a setting in which Information Feedback Regularity ensures equilibrium uniqueness, and where an equilibrium may not exist at all if the regularity condition does not hold.

5.1.1 A Beauty Contest Model

Agent i chooses the price $p_{i,t}$ based on the exogenous signal $s_{i,t}$ and their expectation of the average signal $\overline{p}_t$:

$$p_{i,t} = \varphi s_{i,t} + \alpha E_{i,t}[\overline{p}_t] \tag{26}$$

where $\varphi > 0$ and $\alpha > 0$.

Agents receive the vector of signals:

$$\begin{pmatrix} z_{i,t} \\ s_{i,t} \end{pmatrix} = \begin{pmatrix} \overline{p}_t + \sigma_u u_{i,t} \\ \theta_t + \sigma_v v_{i,t} \end{pmatrix}$$

with θ_t, $u_{i,t}$ and $v_{i,t}$ all distributed $\sim N(0,1)$.

5.1.2 Representation in the General Macroeconomic Framework

This model is structured as a nowcasting problem, while the general framework is structured as a forecasting problem. To map static beauty contests into the general

[25]For a more sophisticated application, see Adams (2023), which studies optimal policy in a model with endogenous signals and capital.

model, we will introduce an additional signal $k_t = p_{t-1}$ which reveals the previous period's average action. Thus agents' signal process is given by

$$A_{i,t} \equiv \begin{pmatrix} z_{i,t} \\ s_{i,t} \\ k_t \end{pmatrix}$$

which in operator form is

$$A_{i,t} = S(L)\varepsilon_{i,t} = \begin{pmatrix} 0 & \sigma_u & 0 \\ 1 & 0 & \sigma_v \\ 0 & 0 & 0 \end{pmatrix} \varepsilon_{i,t} + \begin{pmatrix} 1 \\ 0 \\ L \end{pmatrix} p_t$$

where $\varepsilon_{i,t} = \begin{pmatrix} \theta_t \\ u_{i,t} \\ v_{i,t} \end{pmatrix}$. The operator $G(L) = \begin{pmatrix} 1 \\ 0 \\ L \end{pmatrix}$.

With this signal structure, the equilibrium condition (26) rewritten in the general model form is

$$0 = E_{i,t}\left[-p_{i,t} + \begin{pmatrix} 0 & \varphi & 0 \end{pmatrix} A_{i,t} + \begin{pmatrix} 0 & 0 & \alpha \end{pmatrix} A_{i,t+1}\right]$$

Thus the matrices mapping to the general representation (1) are given by

$$B_{X0} = -1 \quad B_{X1} = 0 \quad B_{A0} = \begin{pmatrix} 0 & \varphi & 0 \end{pmatrix} \quad B_{A1} = \begin{pmatrix} 0 & 0 & \alpha \end{pmatrix}$$

with the policy function for $p_{i,t} = X(L)\varepsilon_{i,t}$:

$$X(L) = \left[\left(B_{A0} + B_{A1}L^{-1}\right) A(L)\right]_+$$

which implies $\Theta(L)\Xi(L) = I$.

For the beauty contest, the signal operator $\mathcal{B}(S) = S_X(L) + G(L)X(L)W(L)P_G$ is

$$\mathcal{B}(S) = \begin{pmatrix} 0 & \sigma_u & 0 \\ 1 & 0 & \sigma_v \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ L \end{pmatrix} \left[\begin{pmatrix} 0 & \varphi & \alpha L^{-1} \end{pmatrix} A(L)\right]_+ W(L)$$

When is Information Feedback Regularity satisfied? The non-causal IFR operator in Condition 1 is $G\Theta\Xi(B_{A1}L^{-1} + B_{A0})$. In the beauty contest $G = \begin{pmatrix} 1 \\ 0 \\ L \end{pmatrix}$, $\Theta\Xi = I$, and $(B_{A1}L^{-1}+B_{A0}) = \begin{pmatrix} 0 & \varphi & \alpha L^{-1} \end{pmatrix}$ so Information Feedback Regularity is satisfied if $\left\| \begin{pmatrix} 0 & \varphi & \alpha L^{-1} \\ 0 & 0 & 0 \\ 0 & L\varphi & \alpha \end{pmatrix} \right\| < 1$.

When is this true? In the following section the norm is calculated numerically, as usual. But sometimes helpful analytical bounds can be found, because norms are bounded below by the norms of rows and columns. Thus Proposition 6 gives some necessary conditions.

Proposition 6 *Information Feedback Regularity is satisfied in the beauty contest model only if $\varphi < \frac{1}{\sqrt{2}}$, $\alpha < \frac{1}{\sqrt{2}}$, and $\varphi^2 + \alpha^2 < 1$.*

Proof: Appendix A.11.2

The proof serves as an example of how to construct the block Toeplitz representation of a non-causal operator, how to lower bound its norm analytically, and how to translate it to a necessary condition on the parameters of an economic model.

The upper bound on α implies that for IFR to hold, the beauty contest cannot feature too much strategic complementarity. If agents put too much weight on being close to the average forecast, the information feedback will be too strong. As the next section shows, this can allow for multiple equilibria, or no equilibrium at all.

5.1.3 Beauty Contest Equilibrium Properties

The beauty contest with endogenous information is an example of when checking the Information Feedback Regularity condition is useful in practice, because this model can have multiple solutions. Theorem 6 guarantees that there can be only one signal-stable solution. In this model, the regularity condition is even stronger in practice than in theory: it rules out multiple equilibria entirely.

The beauty contest has known analytical solutions. θ_t is the only aggregate shock, so let $\overline{p}_t = b\theta_t$ with b to be found.

How do agents nowcast $\overline{p}_t$? They receive two noisy signals of θ_t: $s_{i,t} = \theta_t + \sigma_v v_{i,t}$ with precision $\tau_v \equiv \frac{1}{\sigma_v^2}$ and $\frac{z_{i,t}}{b} = \theta_t + \frac{\sigma_u u_{i,t}}{b}$ with precision $\tau_u b^2 \equiv \frac{b^2}{\sigma_u^2}$. By Lemma 12 their expectation of θ_t is

$$E_{i,t}\left[\theta_t\right] = \frac{\tau_v s_{i,t} + \tau_u b^2 \frac{z_{i,t}}{b}}{1 + \tau_v + \tau_u b^2}$$

They choose the price $p_{i,t}$ by

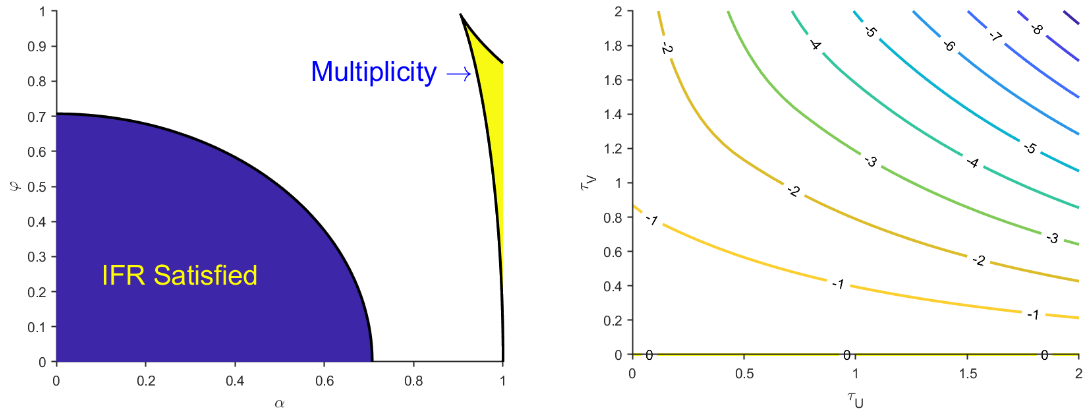
$$\begin{aligned} p_{i,t} &= \varphi s_{i,t} + \alpha E_{i,t}[b\theta_t] \\ &= \varphi s_{i,t} + \alpha b \frac{\tau_v s_{i,t} + \tau_u b^2 \frac{z_{i,t}}{b}}{1 + \tau_v + \tau_u b^2} \end{aligned}$$

The noise shocks $u_{i,t}$ and $v_{i,t}$ are mean zero across agents, so the average price is given by

$$\begin{aligned} \overline{p}_t &= \varphi\theta_t + \alpha E_{i,t}[b\theta_t] \\ &= \varphi\theta_t + \alpha b \frac{\tau_v \theta_t + \tau_u b^2 \theta_t}{1 + \tau_v + \tau_u b^2} \end{aligned}$$

The conjecture $\overline{p}_t = b\theta_t$ implies a single equation for b:

$$b = f(b) \equiv \varphi + \alpha b \frac{\tau_v + \tau_u b^2}{1 + \tau_v + \tau_u b^2} \tag{27}$$



(a) Multiplicity and Regularity Regions (b) Maximum Discriminants under IFR

Figure 2: Equilibrium Uniqueness in the Beauty Contest with Endogenous Signals

The left panel plots the subsets of the parameter space where the beauty contest model has a unique or multiple solutions, with precision parameters $\tau_u = 0.3$ and $\tau_v = 0.15$. The blue region identifies the parameters for which Information Feedback Regularity holds. The right panel shows that the disjointedness of the yellow and blue subspaces are not specific to the parameterization, by plotting the maximum discriminant achievable by choosing α and φ, for each combination of τ_u and τ_v. A negative discriminant implies the model has a unique solution.

which has a cubic representation:

$$b^3\tau_u(1-\alpha) - b^2\tau_u\varphi + b(1+\tau_v(1-\alpha)) - \varphi(1+\tau_v) = 0 \tag{28}$$

There is a unique b that satisfies this equation if and only if the discriminant of the cubic is strictly negative. In the edge case where the discriminant is exactly zero, then there are two unique values for b that satisfy the cubic, corresponding to one signal-stable and one signal-unstable equilibrium. Information Feedback Regularity guarantees that one of these two cases must be true. As an example, consider the case where $\tau_u = 0.3$ and $\tau_v = 0.15$. Figure 2 panel (a) plots the determinacy region under this calibration for different choices of α and φ. When α and φ are near 1, there are multiple solutions.[26] Otherwise, there is a unique solution.

The region where Information Feedback Regularity is satisfied is unconnected: numerically, the regularity condition guarantees a unique solution. And the Sufficient Idiosyncrasy Condition implies IFR, so the region where SIC holds (a subset of the blue region) must also be disjoint from the multiplicity region (Theorem 7). What of the intermediate region without multiplicity but where IFR fails? In this region of the parameter space, there is a unique solution but it must be *signal-unstable* (Theorem 3).

[26] IFR fails in these cases, so no solution is signal stable by Theorem 3. However, two solutions are "exponentially" stable (Appendix B).

In this model, an equilibrium is signal-stable if $\overline{p}_t$ is not too sensitive to perturbations to the signals in any direction. This means that it is not enough to assess stability by comparing the sensitivity of value b alone: if $f'(b) < 1$ (for $f(b)$ defined as in equation (27)), it does not imply signal-stability. $f'(b) < 1$ only implies that $\overline{p}_t$ is stable with respect to deviations in the θ_t dimension of $z_{i,t}$. Signal-stability in this paper is a stronger condition. If IFR fails in the beauty contest model, then there exists a perturbation that is linear in $z_{i,t}$ and $s_{i,t}$ that changes b by more than the perturbation size.

The uniqueness result is independent of the model's calibration. Panel (b) demonstrates that this unconnectedness holds for any choice of τ_u and τ_v. The contours in panel (b) represent the maximum discriminant such that Information Feedback Regularity holds for any pair of $\varphi \in (0, 1)$ and $\alpha \in (0, 1)$. All interior points in this space are negative, so the regularity condition only holds when there is a unique signal-stable equilibrium.

5.2 The Singleton Model

This section studies the asset pricing model considered in Singleton (1987) and solved in Nimark (2017). This model serves as an example for using the general structure in Section 2 and evaluating the regularity condition analytically. Second, this model is a widely-used standard that is simple enough to be well-understood but interesting enough to present computational challenges by featuring endogenous signals and infinite-order dynamics. This lets the model serve as a check of the computational accuracy of Signal Operator Iteration. The following environment uses Nimark's structure and notation.

5.2.1 Singleton's Model of Asset Pricing with Dispersed Information

Agents receive an exogenous signal $z_{i,t} = \theta_t + \sigma_\eta \eta_{i,t}$ where θ_t is an aggregate fundamental and $\eta_{i,t}$ is idiosyncratic white noise. Agents also observe the market-clearing price p_t, which satisfies

$$p_t = -(\theta_t + \sigma_\epsilon \epsilon_t) + \beta f_t$$

where ϵ_t is an aggregate supply shock. θ_t is assumed to be AR(1):

$$\theta_t = \rho \theta_{t-1} + \sigma_u u_t$$

with u_t, ϵ_t, and $\eta_{i,t}$ all standard normal. f_t denotes the average forecast:

$$f_t = \int_i E_{i,t}[p_{t+1}] di$$

and $\beta \in (0, 1)$ is the discount factor.

This model can be neatly put into the general form (1) by letting the individual forecast $f_{i,t}$ be the endogenous variable:

$$X_{i,t} = f_{i,t} \qquad A_{i,t} = \begin{pmatrix} p_t \\ z_{i,t} \end{pmatrix}$$

With this assignment, the agent's equilibrium condition is $f_{i,t} = E_{i,t}[p_{t+1}]$ and the model's matrices are

$$B_{X0} = 1 \quad B_{X1} = 0 \quad B_{A0} = 0 \quad B_{A1} = \begin{pmatrix} -1 & 0 \end{pmatrix}$$

Given the process for $A_{i,t}$, the policy function is

$$\begin{aligned} X(L) &= -[B_{A1}L^{-1}A(L)]_+ \\ &= \begin{pmatrix} 1 & 0 \end{pmatrix} [L^{-1}A(L)]_+ \end{aligned}$$

thus $\Theta(L)\Xi(L) = I$.

The endogenous signal process is

$$S_N(L) = \begin{pmatrix} \beta \\ 0 \end{pmatrix} f_t$$

thus $G(L) = \begin{pmatrix} \beta \\ 0 \end{pmatrix}$

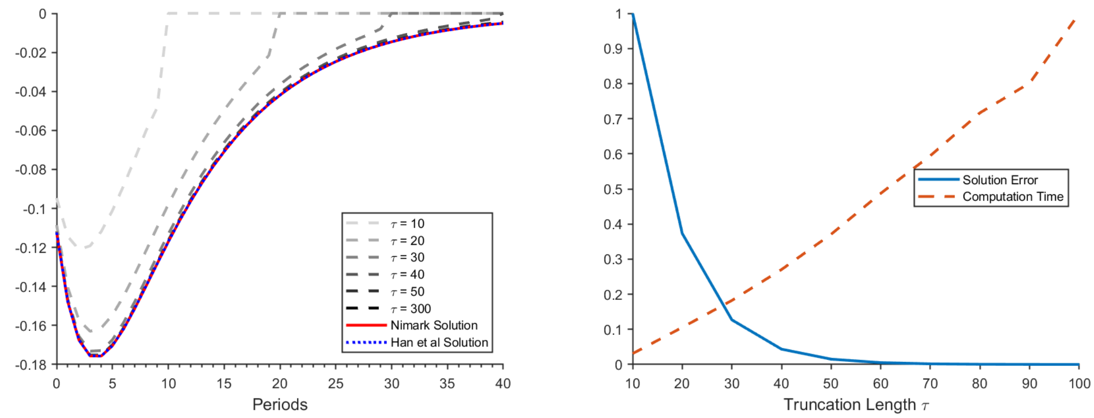
Is Information Feedback Regularity satisfied? Always:

$$\begin{aligned} \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| &= \| \begin{pmatrix} \beta \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \end{pmatrix} L^{-1}\| \\ &= \beta < 1 \end{aligned}$$

Therefore a signal-stable equilibrium is possible (Theorem 3) and there will be at most one (Theorem 6). However, global uniqueness is not guaranteed by any theorem: the Sufficient Idiosyncrasy Condition fails because one signal is entirely aggregate. Instead, global uniqueness can be checked (but not conclusively proven) by searching numerically for alternatives. I have found none.

5.2.2 Singleton Solution

I solve this model in three ways to compare the efficacy and accuracy of Signal Operator Iteration. First, I replicate the original Nimark (2017) solution which uses a Kalman filter and tracks higher order expectations. Second, I apply the Han, Tan, and Wu (2022) method of analytic policy function iteration, which approximates time series with rational polynomials. Third, I apply Signal Operator Iteration with several different truncation orders.



(a) Price Responses to Fundamental Shocks (b) Computation Time and Error

Figure 3: Singleton Model Computation by Truncation Order

The left panel plots the impulse response function in the singleton model for different truncation orders. The right panel plots how the computation time and accuracy depends on the truncation order.

Figure 3 panel (a) plots the impulse response of the asset price to a one standard deviation shock u_t to the fundamental. Each impulse response is calculated in a different way. The solid red line is the solution from Nimark (2017). The dotted blue line uses the algorithm introduced by Han, Tan, and Wu (2022). They overlap almost perfectly. The gray dashed lines correspond to Signal Operator Iterator with different truncation orders. When the truncation order is small (e.g. $\tau = 10$) the fixed point is not a good approximation of the true solution, which has large covariances beyond 10 lags. However, as the truncation order increases, the solutions converge to Nimark's. For $\tau \geq 50$, the impulse response functions are visually indistinguishable.

Increasing the truncation order increases accuracy. But Figure 3 panel (b) demonstrates that there is a computational trade-off. The solid blue line plots to the solution error against the approximation order. For a fixed point $\hat{S}_\tau$ of approximated signal operator $\mathcal{B}_\tau$, I calculate the solution error as $\|\hat{S}_\tau - \hat{S}_{300}\|_S$ (i.e. relative to the solution for a high order approximation). Then I normalize the errors relative to the highest value at $\tau = 10$. As expected (Theorem 2) the solution error decreases as τ increases. The trade-off is that the computation takes longer: the dashed red line plots computation time (normalized relative to $\tau = 100$) when the convergence criterion is $\| \cdot \|_S = 10^{-6}$.

At least in this model, Signal Operator Iteration is not as efficient as the Han, Tan, and Wu (2022) method, who approximate time series with an ARMA process instead of the MA used in my approach. Using the default settings for their algorithm, the model converges approximately $2-3$ times faster than Signal Operator Iteration takes to achieve the same solution error. For computational efficiency and generality, their method is preferred. The advantage of Signal Operator Iteration is its simplicity and known theoretical properties. While most results from Section 3 apply to models independently of how they are solved, the convergence results (Theorems 2 and 5) are specific to Signal Operator Iteration.

5.2.3 Information Feedback Regularity in the Singleton Model

Checking Information Feedback Regularity is useful for practitioners who want to know if they can solve a model. Theorem 3 implies that if IFR is not satisfied, then there cannot exist a signal-stable equilibrium, so a practitioner can find it difficult to calculate a solution. In practice, signal-unstable solutions may sometimes be found even when the feedback is modestly > 1, but large feedbacks are challenging to overcome, especially given that no solution is guaranteed to exist at all. The Singleton model is a useful setting to demonstrate this usefulness, because the regularity condition is controlled by a single parameter: IFR is satisfied if and only if $\beta < 1$.

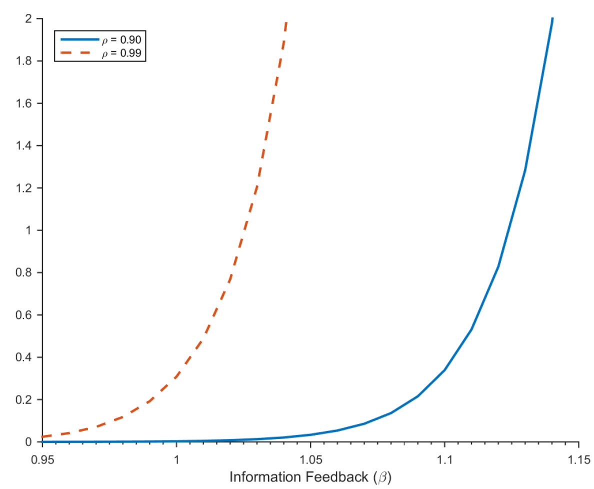


Figure 4: Convergence and Information Feedback Regularity

The figure plots how the signal error in the Singleton model after 50 iterations of SOI depends on the information feedback and autocorrelation. The units are normalized to the signal error of the first guess, which is initialized as the exogenous signal.

Figure 4 demonstrates how the information feedback affects numerical convergence of Signal Operator Iteration by solving the Singleton model for a variety of values of β. In all cases, the initial guess is only the exogenous signal, setting the endogenous component to zero. Then, the y-axis measures the signal error after 50 iterations, relative to the error of the initial guess. The blue line is otherwise the baseline calibration, where $\rho = 0.90$. The baseline model ($\beta = 0.99$) is easily solved with 50 iterations, so the relative error is near zero. And even when $\beta = 1$ and IFR is no longer satisfied, the algorithm still converges, even though the solution is no longer signal-stable. However, when the information feedback increases further, the algorithm no longer converges, and diverges rapidly when $\beta > 1.15$; in this range, the relative error is greater than one, so repeated iterations have increasingly large errors. But IFR controls signal-stability independent of the exogenous signal component, so an alternative exogenous signal may behave differently. This is the case for the red dashed line, which plots the relative errors when $\rho = 0.99$. In this case, signal operators diverge rapidly when the information feedback is only 1.05.

5.3 Unstable Island-Level Problems

Information Feedback Regularity can help clarify why some types of models feature instability that makes them challenging to solve numerically. In the beauty contest model (Section 5.1), instability occurred when the feedback from economic decisions to information was large. Alternatively, instability may occur when the feedback from information to economic decisions is large. In both cases, Information Feedback Regularity fails.

This latter type of instability often occurs in models where agents learn from cross-island asset markets. To demonstrate how this affects the feedback, I study the dispersed information New Keynesian model of Lorenzoni (2009).

5.3.1 The Lorenzoni Model

Lorenzoni (2009) features a New Keynesian model with dispersed information. Agents learn from endogenous signals including cross-island demand, inflation, and nominal interest rates. I adopt the original notation, except islands are denoted by i, and the nominal interest rate is r_t.

The island's problem is characterized by two equations. The first is an Euler equation:

$$0 = E_{i,t}\left[c_{i,t} - c_{i,t+1} + r_t - \overline{\pi}_{i,t+1}\right]$$

where $\overline{\pi}_{i,t+1}$ denotes inflation in island i's consumer goods. Consumption $c_{i,t}$ is an endogenous control, while agents take both the nominal interest rate r_t and $\overline{\pi}_{i,t+1}$ as exogenous. The second equation is an island-specific New Keynesian Phillips curve:

$$0 = E_{i,t}\left[p_{i,t-1} - \left(\lambda(1+\zeta\gamma) + \beta + 1\right)p_{i,t} + \beta p_{i,t+1} + \lambda c_{i,t} + \lambda\left(\overline{p}_{i,t} - (1+\zeta)a_{i,t} + \zeta d_{i,t}\right)\right]$$

where the producer price $p_{i,t}$ is an endogenous control. Agents take the remaining variables as exogenous signals: the input price $\overline{p}_{i,t}$, productivity $a_{i,t}$ and demand $d_{i,t}$. $\beta \in (0,1)$ is the discount factor, $\zeta > 0$ is the Frisch elasticity, $\gamma > 0$ is the elasticity of substitution across islands, and $\lambda \equiv \frac{(1-\theta)(1-\beta\theta)}{\theta} > 0$, where $\theta \in (0,1)$ is the Calvo price adjustment parameter.

This model features unit roots in prices (typical for a New Keynesian model) as well as real variables (because aggregate productivity follows a random walk) so I perform two transformations. First, I define $\tilde{c}_{i,t} \equiv c_{i,t} - a_{i,t}$. Second, I express the New Keynesian Phillips curve in differences, denoted by Δ, with inflations written $\pi_{i,t} = \Delta p_{i,t}$ and $\overline{\pi}_{i,t} = \Delta \overline{p}_{i,t}$. The equations become:

$$0 = E_{i,t}\left[\tilde{c}_{i,t} - \tilde{c}_{i,t+1} + r_t - \Delta a_{i,t+1} - \overline{\pi}_{i,t+1}\right] \tag{29}$$

$$0 = \pi_{i,t-1} - \left(\lambda(1+\zeta\gamma)+1\right)\pi_{i,t} + \beta\Delta E_{i,t}\left[\pi_{i,t+1}\right] + \lambda\Delta\tilde{c}_{i,t} + \lambda\left(\overline{\pi}_{i,t} - \zeta\Delta a_{i,t} + \zeta\Delta d_{i,t}\right) \tag{30}$$

5.3.2 Feedback in the Lorenzoni Model

The full information solution to the model characterized by equations (29) and (30) has no unit roots. But there is a problem: the dispersed information model *still* has a unit root in the Euler equation. This unit root occurs because islands take expected asset returns $E_{i,t}\left[r_t - \overline{\pi}_{i,t+1}\right]$ as exogenous. This is exactly the same problem encountered in linearized small open economy models, where additional assumptions need to be made to introduce a stabilizing force to the Euler equation.

With one unit eigenvalue, the conditions for Theorem 1 are not satisfied. The operators $\Theta(L)$ and $\Xi(L)$ can still be constructed from the recursive equations, but the entries in $\Xi(L)$ are not square-summable. As a result, $\|\Xi\| = \infty$ and Information Feedback Regularity will not be satisfied.

How does the feedback manifest? Small changes to an island's information process have large effects on the island's decision-making. Another consequence of this feedback is small differences in islands' realized signals can have permanent effects on island consumption.

If IFR is not satisfied, then there are no signal-stable equilibria (Theorem 3). But that does not mean that the model does not work. The solution method in Lorenzoni (2009) converges for some calibrations. In particular, the numerical properties are known to be sensitive to the calibrated shock variances. This makes sense: in most dispersed information models, sending some shock variances to zero or infinity recovers full or at least common information. So equilibrium properties should be sensitive to the calibration. But Information Feedback Regularity does not capture this sensitivity. The variances affect the exogenous signal process S_X, while IFR describes the feedback that generates the endogenous signal S_N.

6 Conclusion

This paper introduced a new method for representing and solving a general class of macroeconomic models with endogenous information. I introduced the Information Feedback Regularity condition, which is necessary for an equilibrium to be signal-stable. Signal-stable equilibria are locally unique, and can be approximated arbitrarily well if they are infinite dimensional. Then, I proved that a signal-stable equilibrium must be the globally unique signal-stable equilibrium. The Sufficient Idiosyncrasy Condition ensures that all equilibria are signal-stable, therefore the condition guarantees global uniqueness. To demonstrate these results, I applied the signal operator approach to a variety of simple examples from the literature.

Endogenous information may prove valuable for many applications. Macroeconomic models with information frictions that previously relied on exogenous noise, or that made approximations to the information structure, can now be modeled with fully endogenous signals. Such models can be used to answer questions that were impossible when information was exogenous. How can a policymaker influence expectations by affecting endogenous variables? What is the optimal monetary policy in such an environment when additional frictions and complexities are introduced? What about fiscal stabilization or financial regulation? A wide range of policies that affect asset prices, inflation, unemployment, or other endogenous quantities from which agents might draw information can now be more easily addressed. When economists begin projects tackling these questions, they can easily evaluate the Information Feedback Regularity of their models, and draw useful conclusions about equilibrium properties and the behavior of solution algorithms such as Signal Operator Iteration.

References

ACHARYA, S. (2013): "Dispersed beliefs and aggregate demand management," *University of Maryland mimeo.*

ADAMS, J. J. (2022): "Firestorm: Multiplicity in Models with Full Information," *University of Florida mimeo.*

——— (2023): "Moderating noise-driven macroeconomic fluctuations under dispersed information," *Journal of Economic Dynamics and Control*, 156, 104752.

ADAMS, J. J., AND E. ROJAS (2024): "Household Consumption and Dispersed Information," *Journal of Monetary Economics*, 147, 103592.

ALLEN, F., S. MORRIS, AND H. S. SHIN (2006): "Beauty Contests and Iterated Expectations in Asset Markets," *The Review of Financial Studies*, 19(3), 719–752.

ANGELETOS, G.-M., C. HELLWIG, AND A. PAVAN (2007): "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks," *Econometrica*, 75(3), 711–756.

ANGELETOS, G.-M., AND J. LA'O (2010): "Noisy Business Cycles," *NBER Macroeconomics Annual*, 24, 319–378.

ANGELETOS, G.-M., AND C. LIAN (2020): "Confidence and the Propagation of Demand Shocks," Discussion Paper w27702, National Bureau of Economic Research.

ANGELETOS, G.-M., AND I. WERNING (2006): "Crises and Prices: Information Aggregation, Multiplicity, and Volatility," *American Economic Review*, 96(5), 1720–1736.

BLANCHARD, O. J., AND C. M. KAHN (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48(5), 1305–1311.

BROWDER, F. E. (1954): "Covering spaces, fibre spaces, and local homeomorphisms," *Duke Math. J.*, 21(1), 329–336.

BÖTTCHER, A., AND B. SILBERMANN (2012): *Introduction to Large Truncated Toeplitz Matrices*. Springer Science & Business Media, Google-Books-ID: 64LTBwAAQBAJ.

——— (2013): *Analysis of Toeplitz Operators*. Springer Science & Business Media.

CAINES, P. E., AND L. GERENCSER (1991): "A Simple Proof for a Spectral Factorization Theorem," *IMA Journal of Mathematical Control and Information*, 8(1), 39–44.

CARTAN, H. (1971): *Calcul Differentiel*. Hermann.

CHAHROUR, R., AND G. GABALLO (2020): "Learning from House Prices: Amplification and Business Fluctuations," *The Review of Economic Studies*.

CONWAY, J. B. (2007): *A course in functional analysis*, vol. 96. Springer Science & Business Media, 2nd edn.

FRAZHO, A. E., AND W. BHOSRI (2010): "Toeplitz and Laurent Operators," in *An Operator Perspective on Signals and Systems*, ed. by A. E. Frazho, and W. Bhosri, Operator Theory: Advances and Applications, pp. 23–40. Birkhäuser Basel, Basel.

GARAB, B., M. PITUK, AND C. PÖTZSCHE (2020): "Linearized stability in the context of an example by Rodrigues and Solà-Morales," *Journal of Differential Equations*, 269(11), 9838–9845.

GRAHAM, L., AND S. WRIGHT (2010): "Information, heterogeneity and market incompleteness," *Journal of Monetary Economics*, 57(2), 164–174.

GROSSMAN, S. (1976): "On the Efficiency of Competitive Stock Markets Where Trades Have Diverse Information," *The Journal of Finance*, 31(2), 573–585.

GUTÚ, O. (2017): "On global inverse theorems," *Topological Methods in Nonlinear Analysis*, 49(2), 401–444.

HAN, Z., F. TAN, AND J. WU (2022): "Analytic policy function iteration," *Journal of Economic Theory*, 200, 105395.

HANSEN, L. P., AND T. J. SARGENT (1981): "A note on Wiener-Kolmogorov prediction formulas for rational expectations models," *Economics Letters*, 8(3), 255–260.

HELLWIG, C., A. MUKHERJI, AND A. TSYVINSKI (2006): "Self-Fulfilling Currency Crises: The Role of Interest Rates," *American Economic Review*, 96(5), 1769–1787.

HENDERSON, H. V., AND S. R. SEARLE (1981): "The vec-permutation matrix, the vec operator and Kronecker products: a review," *Linear and Multilinear Algebra*, 9(4), 271–288.

HOLTZMAN, J. M. (1968): "The Use of the Contraction Mapping Theorem with Derivatives in a Banach Space," *Quarterly of Applied Mathematics*, 26(3), 462–465.

HUO, Z., AND M. PEDRONI (2020): "A Single-Judge Solution to Beauty Contests," *American Economic Review*.

HUO, Z., AND N. TAKAYAMA (2015): "Higher order beliefs, confidence, and business cycles," *Working Paper*.

——— (2023): "Rational Expectations Models with Higher-Order Beliefs," *Available at SSRN 3873663*.

JURY, M. (2023): "A Note on Homotopies of Rational Matrix Inner Functions," *manuscript, University of Florida*.

KANTOROVICH, L., AND G. AKILOV (1959): "Functional Analysis in Normed Spaces [in Russian]," *Moscow, Fizmatgiz*.

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

Katriel, G. (1994): “Mountain pass theorems and global homeomorphism theorems,” *Annales de l'Institut Henri Poincaré C, Analyse non linéaire*, 11(2), 189–209.

Keynes, J. M. (1936): *The General Theory of Employment, Interest, and Money*. Macmillan.

Klein, P. (2000): “Using the generalized Schur form to solve a multivariate linear rational expectations model,” *Journal of Economic Dynamics and Control*, 24(10), 1405–1423.

Levine, P., J. Pearlman, S. H. Wright, and B. Yang (2024): “Imperfect Information and Hidden Dynamics,” .

Lorenzoni, G. (2009): “A Theory of Demand Shocks,” *American Economic Review*, 99(5), 2050–2084.

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

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. (2016): “Signaling Effects of Monetary Policy,” *The Review of Economic Studies*.

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

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

——— (2017): “Dynamic Higher Order Expectations,” *Working Paper*.

Rondina, G., and T. Walker (2015): “Dispersed Information and Confounding Dynamics,” *Working Paper*.

Rondina, G., and T. B. Walker (2021): “Confounding Dynamics,” *Journal of Economic Theory*, 196, 105251.

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

Sims, C. A. (2003): “Implications of rational inattention,” *Journal of Monetary Economics*, 50(3), 665–690.

Singleton, K. J. (1987): “Asset prices in a time-series model with disparately informed, competitive traders,” .

Strohmer, T. (2002): "Four short stories about Toeplitz matrix calculations," *Linear Algebra and its Applications*, 343-344, 321–344.

Townsend, R. M. (1983): "Forecasting the Forecasts of Others," *Journal of Political Economy*, 91(4), 546–588.

Uhlig, H. (1995): "A toolkit for analyzing nonlinear dynamic stochastic models easily," *Federal Reserve Bank of Minneapolis Discussion Papers.*

Venkateswaran, V. (2014): "Heterogeneous Information and Labor Market Fluctuations," SSRN Scholarly Paper ID 2687561, Social Science Research Network, Rochester, NY.

Woodford, M. (2003): "The Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps*, ed. by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford.

A Proofs

This appendix contains the proofs of all Theorems, as well as intermediate results.

A.1 Deriving the Policy Function

To simplify notation, define the matrices $\tilde{A}_k$ by

$$\tilde{A}_k \equiv \begin{cases} 0 & k < 0 \\ Q^*(B_{A1}A_{k+1} + B_{A0}A_k) & k \geq 0 \end{cases} \tag{31}$$

with associated lag operator polynomial $\tilde{A}(L) \equiv \sum_{k=-\infty}^{\infty} \tilde{A}_k L^k$. The remaining matrices in the following proof are defined in Section 2.1.

Proof of Theorem 1. The equilibrium conditions (5) must hold for all realizations of the shocks, so it's possible to collect terms, restricting the values of the matrices $\{X_j\}_{j=0}^{\infty}$. This implies a recursive equation for $j \geq 1$:

$$0 = B_{X0}X_j + B_{X1}X_{j+1} + B_{A0}A_j + B_{A1}A_{j+1} \tag{32}$$

Left multiply by Q^*, substitute with $\tilde{A}$, and rearrange to get

$$T_1 Z X_{j+1} = -T_0 Z X_j - \tilde{A}_j \tag{33}$$

The recursive relationship can now be separated into a stable recursive equation and an unstable recursive equation. Let $(ZX)_{C,j}$ and $(Z\tilde{A})_{C,j}$ denote the last n_C rows of $(ZX)_j$ and $\tilde{A}_j$ respectively. Then the unstable recursive equation is

$$T_{1,CC}(ZX)_{C,j+1} = -T_{0,CC}(ZX)_{C,j} - \tilde{A}_{C,j} \tag{34}$$

And where $(ZX)_{S,j}$ and $\tilde{A}_{S,j}$ denote the corresponding first n_S rows, the stable recursive equation is

$$T_{1,SS}(ZX)_{S,j+1} + T_{1,SC}(ZX)_{C,j+1} = -T_{0,SS}(ZX)_{S,j} - T_{0,SC}(ZX)_{C,j} - \tilde{A}_{S,j} \tag{35}$$

Because $T_{0,CC}^{-1}T_{1,CC}$ has all eigenvalues inside the unit circle, the unstable recursive equation (34) allows $(ZX)_{C,j}$ to be expressed as

$$(ZX)_{C,j} = -\sum_{k=0}^{\infty} \left(-T_{0,CC}^{-1}T_{1,CC}\right)^k T_{0,CC}^{-1}\tilde{A}_{C,j+k} \quad \forall j \geq 0 \tag{36}$$

which in lag operator notation is

$$(ZX)_C(L) = \left[-\left(1 + T_{0,CC}^{-1}T_{1,CC}L^{-1}\right)^{-1} T_{0,CC}^{-1}\tilde{A}_C(L)\right]_+$$

Similarly, the stable recursive equation (35) implies the sum $\forall j > 0$

$$(ZX)_{S,j} = \\ -\sum_{k=1}^{j} \left(-T_{1,SS}^{-1}T_{0,SS}\right)^{k-1} T_{1,SS}^{-1} \left(\tilde{A}_{S,j-k} + T_{0,SC}(ZX)_{C,j-k} + T_{1,SC}(ZX)_{C,j+1-k}\right) \\ + \left(-T_{1,SS}^{-1}T_{0,SS}\right)^{j} (ZX)_{S,0} \quad (37)$$

which in lag operator notation is

$$(ZX)_S(L) = \left(I + T_{1,SS}^{-1}T_{0,SS}L\right)^{-1} \\ \left((ZX)_{S,0} - T_{1,SS}^{-1}L\left(\tilde{A}_S(L) + T_{0,SC}(ZX)_C(L) + T_{1,SC}\left[L^{-1}(ZX)_C(L)\right]_+\right)\right)$$

then use $L\left[L^{-1}(ZX)_C\right]_+ = (ZX)_C - (ZX)_{C,0}$ to eliminate the annihilation operator

$$(ZX)_S = \left(I + T_{1,SS}^{-1}T_{0,SS}L\right)^{-1} \\ \left((ZX)_{S,0} + T_{1,SS}^{-1}T_{1,SC}(ZX)_{C,0} - T_{1,SS}^{-1}L\tilde{A}_S(L) - T_{1,SS}^{-1}\left(T_{1,SC} + T_{0,SC}L\right)(ZX)_C(L)\right)$$

Equation (36) determines $(ZX)_{C,j}$ for all $j \geq 0$, but equation (37) only determines $(ZX)_{S,j}$ for $j > 0$. Instead, $(ZX)_{S,0}$ is determined by the restriction that n_S state variables are predetermined. To calculate the initial matrix X_0, relate it to the transformed ZX_0 by

$$\begin{pmatrix} (ZX)_{S,0} \\ (ZX)_{C,0} \end{pmatrix} = \begin{pmatrix} Z_{SS} & Z_{SC} \\ Z_{CS} & Z_{CC} \end{pmatrix} \begin{pmatrix} X_{S,0} \\ X_{C,0} \end{pmatrix}$$

where $X_{S,0}$ are the entries corresponding to the state variables (the first n_S entries in X_0) and $X_{C,0}$ correspond to the controls. The restriction $X_{S,0} = 0$ implies

$$(ZX)_{C,0} = Z_{CC}X_{C,0}$$

Z_{CC} is full rank by assumption, so $(ZX)_{S,0}$ can be found by

$$(ZX)_{S,0} = Z_{SC}Z_{CC}^{-1}(ZX)_{C,0} \quad (38)$$

$(ZX)_{C,0}$ is the forecast error $(ZX)_C - L\left[L^{-1}(ZX)_C\right]_+$ so it can be written as

$$(ZX)_{C,0} = \left[-\left(1 + T_{0,CC}^{-1}T_{1,CC}L^{-1}\right)^{-1} T_{0,CC}^{-1}\tilde{A}_C\right]_+ + L\left[L^{-1}\left(1 + T_{0,CC}^{-1}T_{1,CC}L^{-1}\right)^{-1} T_{0,CC}^{-1}\tilde{A}_C\right]_+ \quad (39)$$

Equations (36), (37), (38), and (39) can be expressed as a single equation with lag operator polynomials:

$$\begin{pmatrix} I & 0 & -I \\ -B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} + T_{1,SS}^{-1}T_{1,SC}\right) & I & B_S(L)^{-1}T_{1,SS}^{-1}\left(T_{1,SC} + T_{0,SC}L\right) \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} (ZX)_{C,0} \\ (ZX)_S \\ (ZX)_C \end{pmatrix} =$$
$$\begin{pmatrix} -L & 0 & 0 \\ 0 & -B_S(L)^{-1}T_{1,SS}^{-1}L & 0 \\ 0 & 0 & I \end{pmatrix} \left[\begin{pmatrix} 0 & -L^{-1}B_C(L)^{-1}T_{0,CC}^{-1} \\ I & 0 \\ 0 & -B_C(L)^{-1}T_{0,CC}^{-1} \end{pmatrix} \tilde{A}(L) \right]_+$$

where $B_S(L) \equiv (I + T_{1,SS}^{-1}T_{0,SS}L)$ and $B_C(L) \equiv (I + T_{0,CC}^{-1}T_{1,CC}L^{-1})$ The left operator is easily inverted:

$$\begin{pmatrix} (ZX)_{C,0} \\ (ZX)_S \\ (ZX)_C \end{pmatrix} = \begin{pmatrix} I & 0 & I \\ B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} + T_{1,SS}^{-1}T_{1,SC}\right) & I & B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} - T_{1,SS}^{-1}T_{0,SC}L\right) \\ 0 & 0 & I \end{pmatrix}$$
$$\begin{pmatrix} -L & 0 & 0 \\ 0 & -B_S(L)^{-1}T_{1,SS}^{-1}L & 0 \\ 0 & 0 & I \end{pmatrix} \left[\begin{pmatrix} 0 & -L^{-1}B_C(L)^{-1}T_{0,CC}^{-1} \\ I & 0 \\ 0 & -B_C(L)^{-1}T_{0,CC}^{-1} \end{pmatrix} \tilde{A}(L) \right]_+$$

Select the second and third block rows:

$$ZX(L) =$$
$$\begin{pmatrix} -B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} + T_{1,SS}^{-1}T_{1,SC}\right)L & -B_S(L)^{-1}T_{1,SS}^{-1}L & B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} - T_{1,SS}^{-1}T_{0,SC}L\right) \\ 0 & 0 & I \end{pmatrix}$$
$$\left[\begin{pmatrix} 0 & -L^{-1}B_C(L)^{-1}T_{0,CC}^{-1} \\ I & 0 \\ 0 & -B_C(L)^{-1}T_{0,CC}^{-1} \end{pmatrix} \tilde{A}(L) \right]_+$$

and left multiplying by Z^* yields

$$X(L) =$$
$$Z^* \begin{pmatrix} -B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} + T_{1,SS}^{-1}T_{1,SC}\right)L & -B_S(L)^{-1}T_{1,SS}^{-1}L & B_S(L)^{-1}\left(Z_{SC}Z_{CC}^{-1} - T_{1,SS}^{-1}T_{0,SC}L\right) \\ 0 & 0 & I \end{pmatrix}$$
$$\left[\begin{pmatrix} 0 & -L^{-1}B_C(L)^{-1}T_{0,CC}^{-1} \\ I & 0 \\ 0 & -B_C(L)^{-1}T_{0,CC}^{-1} \end{pmatrix} Q^*Q\tilde{A}(L) \right]_+$$

which uses that Z and Q are unitary.

Substituting in $\Theta(L)$ and $\Xi(L)$ gives

$$X(L) = \Theta(L)[\Xi(L)Q\tilde{A}(L)]_+$$

and substituting with the definition $\tilde{A}(L) = [Q^* (B_{A1}L^{-1} + B_{A0}) A(L)]_+$ gives

$$X(L) = \Theta(L) \left[\Xi(L) \left[\left(B_{A1}L^{-1} + B_{A0}\right) A(L)\right]_+\right]_+$$

$\Xi(L)$ has no causal terms beyond χ_0, so $\left[\Xi(L) \left[B_{A1}L^{-1}\right]_+\right]_+ = \left[\Xi(L)B_{A1}L^{-1}\right]_+$, and eliminating the inner annihilator completes the proof:

$$X(L) = \Theta(L) \left[\Xi(L) \left(B_{A1}L^{-1} + B_{A0}\right) A(L)\right]_+$$

■

A.2 The Signal Space

Lower case variables denote infinite square summable vectors, e.g. $y_i \in \ell^2$. Column vectors y_i are indexed by $i = 1, ...m$; when collected, the m vectors form an $\infty \times m$ block vector y, which is written without a subscript. Upper-case variables denote the corresponding lower triangular block Toeplitz operator with symbol y, e.g. Y has block columns $\{y, Ly, L^2y, ...\}$.

When the blocks are $m \times n$, the lag operator L right-shifts a vector n times. The operator L^{-1} is the left-inverse of L, which left-shifts a vector by the length of its associated block. As usual, Y^* denotes the adjoint of Y; when y_i is real (always the case in this section) y_i^* is the vector transposed.

A.2.1 Norms

Several norms are used. First, if the norm of a vector y_i is written without subscript, then it is the usual ℓ^2 (Euclidean) vector norm:

$$\text{Vector Norm of } y_i: \qquad ||y_i|| = ||Ye_i||$$

where e_i the ith standard basis vector, showing how Y is related to its y_i columns.

Second, an additional "signal norm" is necessary to introduce because of the block structure of the signals' operator representations:

Definition 2 (Signal Norm) *Define the norm $\|Y\|_S$ of a block Toeplitz operator Y as*

$$\|Y\|_S = \sqrt{\sum_{i=1}^{m} \|y_i\|^2}$$

Let e_i denote the ith standard basis column vector; then

$$\|Y\|_S^2 = \sum_{i=1}^{m} \|Ye_i\|^2$$

The squared signal norm is the sum of squared norms of the columns that constitute a block column. Equivalently, the signal norm is the sum of the squared Frobenius norm of each sub-matrix Y_j:

$$\|Y\|_S^2 = \sum_{j=0}^{\infty} \|Y_j\|_F^2 \tag{40}$$

When evaluating a vector, it is simply the ℓ^2 vector norm.

Third, if the norm of a signal is written without subscript, then it is the operator norm of the associated Toeplitz operator:

$$\text{Operator Norm of } Y: \qquad ||Y|| \equiv \sup_{s \, s.t. \, ||s||_S = 1} ||Ys||_S$$

In this paper, the Toeplitz operators map signals to signals. The appropriate vector space is defined below, but the operator norm is easily calculated because it is equivalent to the $\ell^2 \to \ell^2$ operator norm. When considering the norm of an operator X on a signal Y, by definition the norm $\|XY\|_S$ satisfies

$$\|XY\|_S \leq \|X\|\|Y\|_S$$

A.2.2 Definition of the Vector Space

Definition 3 (Signal Space) *Define $\mathcal{S}_{m_A, m_\varepsilon}$ as the set of $m_A \times m_\varepsilon$ lower block triangular operators with finite $\|\cdot\|_S$ norm.*

$\mathcal{S}_{m_A, m_\varepsilon}$ is the Banach space of causal signals, with norm $\|\cdot\|_S$ as the distance metric. A property of this space is that it is closed under addition when the blocks are the same size, and it is closed under multiplication when the block dimensions agree.

A.3 Radius Lemmas

This section proves two lemmas that bound the norms of endogenous signals.

Lemma 1 *If Information Feedback Regularity holds, then all fixed points satisfying $\hat{S} = \mathcal{B}(\hat{S})$ have signal norms*

$$\|\hat{S}\|_S \leq R_S$$

where $R_S = \frac{\|S_X\|_S}{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|}$.

Proof. By equation (20), a fixed point with Wold representation $\hat{S} = AW$ satisfies

$$\hat{S} = S_X + G\Theta \left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ WP_G$$

By the triangle inequality:

$$\|\hat{S}\|_S \leq \|S_X\|_S + \|G\Theta \left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ WP_G\|_S$$

$$\leq \|S_X\|_S + \|G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ W\|_S$$

which uses $\|P_G\|_S = 1$ because P_G is a projection. Next, let $C_W C_W' = \Sigma_W$ denote the Cholesky decomposition of the variance of white noise innovations. As a result, $(C_W^{-1}W)^\intercal$ is an isometry, because $W^{\intercal *} W^\intercal = \Sigma_W$,[27] which implies

$$= \|S_X\|_S + \|G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ C_W C_W^{-1} W\|_S$$

$$= \|S_X\|_S + \|G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})A\right]_+ C_W\|_S$$

because in this case the isometry is a change of basis that does not affect the norm. C_W is a block-diagonal operator so it passes through the annihilator; the inequality becomes

$$\|\hat{S}\|_S \leq \|S_X\|_S + \|G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})AC_W\right]_+ \|_S \tag{41}$$

The signal norm $\|\cdot\|_S$ is just the sum of vector norms in the first column block of an operator. To evaluate the norm in equation (41), let $\tilde{a}$ denote the first block column of AC_W. The block Toeplitz representation of $\Xi(B_{A1}L^{-1} + B_{A0})$ is

$$\begin{pmatrix} \Xi_0 & \Xi_{-1} & \Xi_{-2} & \Xi_{-3} & \dots \\ 0 & \Xi_0 & \Xi_{-1} & \Xi_{-2} & \dots \\ 0 & 0 & \Xi_0 & \Xi_{-1} & \dots \\ 0 & 0 & 0 & \Xi_0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} B_{A0} & B_{A1} & 0 & 0 & \dots \\ 0 & B_{A0} & B_{A1} & 0 & \dots \\ 0 & 0 & B_{A0} & B_{A1} & \dots \\ 0 & 0 & 0 & B_{A0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

$$= \begin{pmatrix} \Xi_0 B_{A0} & \Xi_{-1}B_{A0} + \Xi_0 B_{A1} & \Xi_{-2}B_{A0} + \Xi_1 B_{A1} & \Xi_{-3}B_{A0} + \Xi_{-2}B_{A1} & \dots \\ 0 & \Xi_0 B_{A0} & \Xi_{-1}B_{A0} + \Xi_0 B_{A1} & \Xi_{-2}B_{A0} + \Xi_1 B_{A1} & \dots \\ 0 & 0 & \Xi_0 B_{A0} & \Xi_{-1}B_{A0} + \Xi_0 B_{A1} & \dots \\ 0 & 0 & 0 & \Xi_0 B_{A0} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

which is also block Toeplitz because Ξ and $B_{A1}L^{-1} + B_{A0}$ are both upper block triangular. The product of this block Toeplitz operator and $\tilde{a}$ gives the first block column of $[\Xi(B_{A1}L^{-1} + B_{A0})AC_W]_+$. And $G\Theta$ is causal (i.e. lower block triangular) so premultiplying by the block Toeplitz representation of $G\Theta$ then gives the first block column of $G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})AC_W\right]_+$. As a matter of convention, writing $G\Theta\Xi(B_{A1}L^{-1} + B_{A0})$ followed by any vector implies multiplying the corresponding block Toeplitz operators (G, Θ, Ξ, and $(B_{A1}L^{-1} + B_{A0})$) with that vector, thus

$$\|G\Theta\left[\Xi(B_{A1}L^{-1} + B_{A0})AC_W\right]_+ \|_S = \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\tilde{a}\|_S$$

Continuing the inequality from equation (41):

$$\|\hat{S}\|_S \leq \|S_X\|_S + \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\tilde{a}\|_S$$

[27] $(\cdot)^\intercal$ denotes the block-transpose, which transposes the blocks of a block-Toeplitz operator.

$$\leq \|S_X\|_S + \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| \|\tilde{a}\|_S$$

by definition of the operator norm $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|$

$$= \|S_X\|_S + \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| \|S\|_S$$

because $\|\tilde{a}\|_S = \|AC_W\|_S = \|AC_W C_W^{-1} W\|_S = \|S\|_S$.

Then rearrange the inequality $\|\hat{S}\|_S \leq \|S_X\|_S + \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| \|\hat{S}\|_S$:

$$\|\hat{S}\|_S \leq \frac{\|S_X\|_S}{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|}$$

which is possible if IFR holds so that $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| < 1$. ■

Lemma 2 *If the Information Feedback Regularity Condition holds, and $\hat{S} = \mathcal{B}(\hat{S})$ is a fixed point, then the signal norm of the aggregate component $\hat{S}P_G$ is bounded by*

$$\|\hat{S}P_G\|_S \leq \frac{\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|}$$

where $\vartheta_I = \|G\Theta\| \| \left[\Xi(B_{A1}L^{-1} + B_{A0}) S_X (I - P_G)\right]_+ \|_S$, while the signal norm of the endogenous signal $\hat{S}_N = G\Theta[\Xi(B_{A1}L^{-1} + B_{A0})A]_+ W P_G$ is bounded by

$$\|G\Theta[\Xi(B_{A1}L^{-1} + B_{A0})A]_+ W P_G\|_S \leq R_N$$

where $R_N = \frac{\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\| \|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|}$.

Proof. A fixed point $\hat{S} = AW$ satisfies

$$\hat{S} = S_X + G\Theta \left[\zeta A\right]_+ W P_G$$

where $\Xi(B_{A1}L^{-1} + B_{A0}) = \zeta$. where $P_{\hat{S}}$ denotes projection onto current and past $\hat{S}$ signals. $[\zeta A]_+ W$ is the projection of the noncausal signals $\zeta \hat{S}$ onto current and past $\hat{S}$. To write this projection step as a linear operator, it is necessary to transpose the blocks of the signals without transposing the whole operator, as described in Appendix A.5.1. $\hat{S}^\intercal$ indicates this "block transpose"; the operator is still lower triangular block Toeplitz, but each block is transposed. This approach is useful because Lemma 9 can be applied to bound the effects of certain non-causal operators. As in Appendix A.5.2, $P_{\hat{S}} = \hat{S}^\intercal (\hat{S}^{\intercal *} \hat{S}^\intercal)^{-1} \hat{S}^{\intercal *}$ denotes the projection onto the columns of $\hat{S}^\intercal$.

With this representation, the fixed point satisfies[28]

$$\hat{S} = S_X + G\Theta \left(P_{\hat{S}} [(\zeta \hat{S})^\intercal]_+ \right)^\intercal P_G$$

[28]Note that $(\zeta \hat{S})^\intercal \neq \hat{S}^\intercal \zeta^\intercal$ in general; this separation only holds with equality if ζ is causal.

and P_G is a projection, so the aggregate component of any fixed point satisfies

$$\hat{S}P_G = S_X P_G + G\Theta\left(P_{\hat{S}}[(\zeta\hat{S})^\intercal]_+\right)^\intercal P_G \tag{42}$$

Take the norm and apply the triangle inequality:

$$\|\hat{S}P_G\|_S \le \|S_X P_G\|_S + \|G\Theta\left(P_{\hat{S}}[(\zeta\hat{S})^\intercal]_+\right)^\intercal P_G\|_S \le \|S_X P_G\|_S + \|G\Theta\left(P_{\hat{S}}[(\zeta\hat{S})^\intercal]_+\right)^\intercal\|_S$$

The second inequality holds because P_G is a projection with norm $\|P_G\| = 1$.

By Lemma 9, the inequality becomes

$$\|\hat{S}P_G\|_S \le \|S_X P_G\|_S + \|P_{\hat{S}}\|\left(\|G\Theta\zeta\|\|\hat{S}P_G\|_S + \vartheta_I\right)$$

$P_{\hat{S}}$ is also a projection, so $\|P_{\hat{S}}\| = 1$. If IFR holds, $\|G\Theta\zeta\| < 1$, so the inequality can be rearranged to

$$\|\hat{S}P_G\|_S \le \frac{\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\zeta\|}$$

The definition of ζ proves the first result.

The latter term in equation (42) is the endogenous component; take the signal norm:

$$\begin{aligned}\|G\Theta\left[\zeta A\right]_+ W P_G\|_S &= \|G\Theta\left(P_{\hat{S}}[(\zeta\hat{S})^\intercal]_+\right)^\intercal P_G\|_S \\ &\le \|G\Theta\left(P_{\hat{S}}[(\zeta\hat{S})^\intercal]_+\right)^\intercal\|_S\end{aligned}$$

because P_G is a projection. Apply Lemma 9:

$$\le \|P_{\hat{S}}\|\left(\|G\Theta\zeta\|\|\hat{S}P_G\|_S + \vartheta_I\right) = \|G\Theta\zeta\|\|\hat{S}P_G\|_S + \vartheta_I$$

Then use the first result that $\|\hat{S}P_G\|_S \le \frac{\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\zeta\|}$ to bound

$$\|G\Theta\left[\zeta A\right]_+ W P_G\|_S \le \|G\Theta\zeta\|\frac{\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\zeta\|} + \vartheta_I = \frac{\|G\Theta\zeta\|\|S_X P_G\|_S + \vartheta_I}{1 - \|G\Theta\zeta\|}$$

Again, the definition of ζ proves the second result. ■

A.4 Convergence to an Infinite-Order Fixed Point

Towards the proof of Theorem 2, I next prove that the algorithm $\mathcal{B}_\tau$ approximates $\mathcal{B}$ arbitrarily well for large τ

Lemma 3 *$\mathcal{B}_\tau$ converges to $\mathcal{B}$ pointwise, i.e.*

$$\lim_{\tau\to\infty} \mathcal{B}_\tau(S) = \mathcal{B}(S) \quad \forall S \in \mathcal{S}_{m_A, m_\varepsilon}$$

Proof. Consider any $S \in \mathcal{S}_{m_A,m_\varepsilon}$

$$\begin{aligned}\|\mathcal{B}(S) - \mathcal{B}_\tau(S)\|_S^2 &= \|(S_X + GXWP_G)(I - P_\tau)\|_S^2 \\ &= \|\mathcal{B}(S)(I - P_\tau)\|_S^2 \\ &= \sum_{j=r}^{\infty} \|(\mathcal{B}(S))_j\|_S^2\end{aligned}$$

where $(\mathcal{B}(S))_j$ denotes the jth block of $\mathcal{B}(S)$.

$\mathcal{B}(S)$ is square summable, so for any $\varepsilon > 0$, there exists a K such that $\sum_{j=K}^{\infty} \|(\mathcal{B}(S))_j\|_S^2 < \varepsilon^2$. Therefore $\lim_{\tau\to\infty} \mathcal{B}_\tau(S) = \mathcal{B}(S)$ and $\mathcal{B}_\tau \to \mathcal{B}$ pointwise. ■

Proof of Theorem 2. Lemma 3 says $\mathcal{B}_\tau \to \mathcal{B}$ pointwise, so for any $\frac{\varepsilon}{2} > 0$, there exists a K_1 s.t.

$$\|\mathcal{B}_\tau(S_\tau) - \mathcal{B}(S_\tau)\| < \frac{\varepsilon}{2} \qquad \forall r \geq K_1$$

$\mathcal{B}$ is continuous and $\hat{S}_\tau \to \hat{S}$, so for any $\frac{\varepsilon}{2} > 0$, there exists a K_2 s.t.

$$\|\mathcal{B}(\hat{S}_\tau) - \mathcal{B}(\hat{S})\| < \frac{\varepsilon}{2} \qquad \forall r \geq K_2$$

Therefore:

$$\|\mathcal{B}_\tau(\hat{S}_\tau) - \mathcal{B}(\hat{S}_\tau)\| + \|\mathcal{B}(\hat{S}_\tau) - \mathcal{B}(\hat{S})\| < \varepsilon \qquad \forall r \geq \max(K_1, K_2)$$

Then by the triangle inequality:

$$\|\mathcal{B}_\tau(\hat{S}_\tau) - \mathcal{B}(\hat{S})\| < \varepsilon \qquad \forall r \geq \max(K_1, K_2)$$

so by definition

$$\lim_{\tau\to\infty} \mathcal{B}_\tau(\hat{S}_\tau) = \mathcal{B}(\hat{S})$$

Then substitute with the fixed points, followed by their limit:

$$\lim_{\tau\to\infty} \hat{S}_\tau = \mathcal{B}(\hat{S})$$

$$\hat{S} = \mathcal{B}(\hat{S})$$

■

A.5 The Norm of the Fréchet Derivative

This section introduces some notation useful for characterizing the norm of the Fréchet derivative, states an intermediate lemma, and finally derives the expression.

A.5.1 Additional Notation

Toeplitz operators have an associated Hankel operator. The block Toeplitz operator Y constructed from the block column y has block columns $y, Ly, L^2y, ...$ I denote the associated Hankel operator $H(y)$, which has block columns $y, L^{-1}y, L^{-2}y, ...$, and thus is block-symmetric. If $H(Y)$ is written in terms of a non-Toeplitz operator Y, it is implied that it takes the first block column of Y as its argument.

Some operations are made more difficult by the fact that signals are block Toeplitz operators, rather than regular Toeplitz operators which would otherwise commute for causal signals. One method to resolve this is to permute the blocks into vectors, and apply Kronecker products of operators; this requires some further notation. $bvec(\cdot)$ vectorizes each block of an $(m\times n)$-block operator, producing a $(mn\times 1)$-block Toeplitz operator, by stacking sub-block columns. For example, $bvec(Y)$ is a block Toeplitz operator, and $bvec(Y)e_1$ is its first column, which encodes all of the information of a lower triangular block Toeplitz operator Y into a vector. Therefore the signal and vector norms are related by

$$\|Y\|_S = \|bvec(Y)e_1\|$$

which immediately follows from Definition 2.

Sometimes it is necessary to transpose the blocks of an operator without transposing the entire operator. The block transpose of an operator Y is denoted by $Y^\intercal$.

Let $\mathbf{L}_Y$ and $\mathbf{R}_Y$ denote left and right multiplication of a vectorized operator, such that the original operator is left or right multiplied by Y, respectively. In other words, for block $m\times n$ operator Y and scalar k, the blocks of $\mathbf{L}_Y^k$ and $\mathbf{R}_Y^k$ are given by:

$$\mathbf{L}^k_{Y,ij} = I_k \otimes Y_{ij} \qquad \mathbf{R}^k_{Y,i,j} = Y^*_{i,j} \otimes I_k \tag{43}$$

then conformable operators X (with $k\times m$ blocks) and Y (with $m\times n$ blocks) satisfy

$$bvec(XY) = \mathbf{L}^n_X bvec(Y)$$

Using $\mathbf{R}$ requires more conditions than $\mathbf{L}$. One special case is where X and Y are conformable lower block triangular Toeplitz operators:

$$bvec(XY) = \mathbf{R}^k_Y bvec(X)$$

A second special case is where H and Y are conformable Hankel and lower block triangular Toeplitz operators, respectively:

$$bvec(HY) = \mathbf{R}^k_{Y^{\intercal *}} bvec(H)$$

Hankel and upper triangular Toeplitz operators have a useful relationship in the case where x and y are ordinary vectors: $X^*y = H(y)x$. Property 2 generalizes this to the block case.

Property 2 *For $m \times k$ block vector x and $m \times n$ block vector y:*

$$bvec(X^* y) = \varrho_{k,n} bvec(H(y^\intercal)x)$$

where $\varrho_{k,n}$ is the vec-permutation matrix[29] for $k \times n$ vectorized matrices.

A.5.2 A Lemma for Characterizing the Fréchet Derivative

Consider a lower block triangular signal Toeplitz operator $S \in \mathcal{S}_{m_A, m_\varepsilon}$, and a deviation $S^\Delta \in \mathcal{S}_{m_A, m_\varepsilon}$. Denote the difference $D \equiv S^\Delta - S$. Let $P_S \equiv S^\intercal (S^{\intercal *} S^\intercal)^{-1} S^{\intercal *}$ denote the projection onto the columns of $S^\intercal$, let $M_S \equiv I - P_S$ denote the residual projection and let $S_L^{-\intercal} \equiv (S^{\intercal *} S^\intercal)^{-1} S^{\intercal *}$ denote the left inverse of the signal Toeplitz operator $S^\intercal$.

Lemma 4 *If Φ is a conformable non-causal block Toeplitz operator, then for $i \leq m_\varepsilon$:*

$$(P_{S^\Delta} - P_S)[(\Phi S)^\intercal]_+ e_i = M_S D^\intercal S_L^{-\intercal} [(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} M_S [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S) \tag{44}$$

Proof.

$$(P_{S^\Delta} - P_S)[(\Phi S)^\intercal]_+ e_i = (S^\intercal + D^\intercal)((S^\intercal + D^\intercal)^*(S^\intercal + D^\intercal))^{-1}(S^\intercal + D^\intercal)^*[(\Phi S)^\intercal]_+ e_i - P_S[(\Phi S)^\intercal]_+ e_i$$

$$\begin{aligned} &= D^\intercal (S^{\intercal *} S^\intercal)^{-1} S^* [(\Phi S)^\intercal]_+ e_i + S^\intercal (S^{\intercal *} S^\intercal)^{-1} D^{\intercal *} [(\Phi S)^\intercal]_+ e_i ... \\ &\qquad + S^\intercal ((S^\intercal + D^\intercal)^* (S^\intercal + D^\intercal))^{-1} S^{\intercal *} [(\Phi S)^\intercal]_+ e_i - P_S [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S) \end{aligned}$$

The $o(\|D\|_S)$ term here collects all $D^{\intercal *} D^\intercal$ and $D^\intercal D^{\intercal *}$ terms. Substitute in $S_L^{-\intercal} \equiv (S^{\intercal *} S^\intercal)^{-1} S^{\intercal *}$ to simplify notation:

$$\begin{aligned} &= D^\intercal S_L^{-\intercal} [(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} [(\Phi S)^\intercal]_+ e_i ... \\ &\qquad + S^\intercal ((S^\intercal + D^\intercal)^* (S^\intercal + D^\intercal))^{-1} S^{\intercal *} [(\Phi S)^\intercal]_+ e_i - P_S [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S) \end{aligned}$$

$$\begin{aligned} &= D^\intercal S_L^{-\intercal} [(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} [(\Phi S)^\intercal]_+ e_i ... \\ &\qquad + S^\intercal (S^{\intercal *} S^\intercal + D^{\intercal *} S^\intercal + S^{\intercal *} D^\intercal)^{-1} S^{\intercal *} [(\Phi S)^\intercal]_+ e_i - P_S [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S) \end{aligned}$$

Using $P_S = S^\intercal (S^{\intercal *} S^\intercal)^{-1} S^{\intercal *}$:

$$\begin{aligned} &= D^\intercal S_L^{-\intercal} [(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} [(\Phi S)^\intercal]_+ e_i ... \\ &\qquad + S^\intercal (S^{\intercal *} S^\intercal + D^{\intercal *} S^\intercal + S^{\intercal *} D^\intercal)^{-1} S^{\intercal *} [(\Phi S)^\intercal]_+ e_i ... \\ &- S^\intercal (S^{\intercal *} S^\intercal + D^{\intercal *} S^\intercal + S^{\intercal *} D^\intercal)^{-1} (S^{\intercal *} S^\intercal + D^{\intercal *} S^\intercal + S^{\intercal *} D^\intercal)(S^{\intercal *} S^\intercal)^{-1} S^{\intercal *} [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S) \end{aligned}$$

[29] Henderson and Searle (1981)

$$= D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *}[(\Phi S)^\intercal]_+ e_i ...$$
$$+ S^\intercal (S^{\intercal *}S^\intercal + D^{\intercal *}S^\intercal + S^{\intercal *}D^\intercal)^{-1} \left(I - (S^{\intercal *}S^\intercal + D^{\intercal *}S^\intercal + S^{\intercal *}D^\intercal)(S^{\intercal *}S^\intercal)^{-1}\right) S^{\intercal *}[(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

$$= D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *}[(\Phi S)^\intercal]_+ e_i ...$$
$$- S^\intercal (S^{\intercal *}S^\intercal + D^{\intercal *}S^\intercal + S^{\intercal *}D^\intercal)^{-1} \left(D^{\intercal *}S^\intercal (S^{\intercal *}S^\intercal)^{-1} + S^{\intercal *}D^\intercal (S^{\intercal *}S^\intercal)^{-1}\right) S^{\intercal *}[(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

Subsume some additional terms into $o(\|D\|_S)$:

$$= D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *}[(\Phi S)^\intercal]_+ e_i ...$$
$$- S^\intercal (S^{\intercal *}S^\intercal)^{-1} \left(D^{\intercal *}S^\intercal (S^{\intercal *}S^\intercal)^{-1} + S^{\intercal *}D^\intercal (S^{\intercal *}S^\intercal)^{-1}\right) S^{\intercal *}[(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

$$= D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *}[(\Phi S)^\intercal]_+ e_i ...$$
$$- S^\intercal (S^{\intercal *}S^\intercal)^{-1} D^{\intercal *} P_S [(\Phi S)^\intercal]_+ e_i - P_S D^\intercal (S^{\intercal *}S^\intercal)^{-1} S^{\intercal *}[(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

Substitute $M_S \equiv I - P_S$ and $S_L^{-\intercal} \equiv (S^{\intercal *}S^\intercal)^{-1}S^{\intercal *}$:

$$= D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *}[(\Phi S)^\intercal]_+ e_i ...$$
$$- S_L^{-\intercal *} D^{\intercal *} P_S [(\Phi S)^\intercal]_+ e_i - P_S D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

$$= M_S D^\intercal S_L^{-\intercal}[(\Phi S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} M_S [(\Phi S)^\intercal]_+ e_i + o(\|D\|_S)$$

■

A.5.3 The Fréchet Derivative

To make the equations in the Theorem 8 proof more manageable, define

$$\zeta \equiv \Xi(B_{A1}L^{-1} + B_{A0}) \tag{45}$$

and define $n_\zeta \equiv n + n_s$, the number of row dimensions in the blocks of Ξ.

To simplify the norm itself, I introduce subspace coefficients $\mathbf{Q}_{P_S}$ and $\mathbf{Q}_{M_S}$.

Definition 4 *For nonzero $S \in \mathcal{S}_{m_A,m_\varepsilon}$, define*

$$\mathbf{Q}_{P_S} \equiv \mathbf{R}^{m_\varepsilon}_{\zeta^\intercal}$$

$$\mathbf{Q}_{M_S} \equiv \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} \varrho_{n_\zeta, m_A} \mathbf{L}^{m_A}_{H((M_S[(\zeta S)^\intercal]_+)^\intercal)}$$

Using this notation, Theorem 8 gives the explicit matrix representation of the Fréchet derivative.

Theorem 8 *When a signal S is represented in the vectorized form $bvec(S^\intercal)e_1$, the Fréchet derivative of $\mathcal{B}(S)$ has a matrix representation given by*

$$D_{\mathcal{B}}(S) = \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S} + \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S} \tag{46}$$

Proof. Consider a signal perturbation $D \equiv S^\Delta - S$. The perturbed difference $\mathcal{B}(S^\Delta) - \mathcal{B}(S)$ is given by

$$\mathcal{B}(S^\Delta) - \mathcal{B}(S) = G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right) P_G \tag{47}$$

I will first characterize the interior term $[\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W$.

$[\zeta A]_+ W$ is the projection of the noncausal signals $\zeta AW = \zeta S$ onto current and past W, or equivalently onto current and past S. The block structure of the signals requires some additional care to ensure conformability: in order to project columns of ζS onto the space spanned by lags of S, the blocks must be transposed first. In other words, the columns of the block-transposed operator $([\zeta A]_+ W)^\intercal e_i$ are given by the projection of $[(\zeta S)^\intercal]_+ e_i$ for each $i \leq m$ onto $S^\intercal$.[30] The ith column of the term $[\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W$ for each $i \leq m$ becomes:

$$\begin{aligned}
([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W)^\intercal e_i &= P_{S^\Delta}[(\zeta S^\Delta)^\intercal]_+ e_i - P_S[(\zeta S)^\intercal]_+ e_i \\
&= (P_{S^\Delta} - P_S)[(\zeta S)^\intercal]_+ e_i + P_S[(\zeta D)^\intercal]_+ e_i + o(\|D\|_S) \\
&= M_S D^\intercal S_L^{-\intercal}[(\zeta S)^\intercal]_+ e_i + S_L^{-\intercal *} D^{\intercal *} M_S[(\zeta S)^\intercal]_+ e_i + P_S[(\zeta D)^\intercal]_+ e_i + o(\|D\|_S)
\end{aligned}$$

by Lemma 4. Block vectorize:

$$\begin{aligned}
&bvec(([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W)^\intercal)e_1 = \\
&bvec(M_S D^\intercal S_L^{-\intercal}[(\zeta S)^\intercal]_+)e_1 + bvec(S_L^{-\intercal *} D^{\intercal *} M_S[(\zeta S)^\intercal]_+)e_1 + bvec(P_S[(\zeta D)^\intercal]_+)e_1 + o(\|D\|_S)
\end{aligned}$$

The vector e_1 selects the first column, making the $[\cdot]_+$ operators redundant on the right-hand side because S and D are lower triangular:

$$\begin{aligned}
&bvec(([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W)^\intercal)e_1 = \\
&bvec(M_S D^\intercal S_L^{-\intercal}(\zeta S)^\intercal)e_1 + bvec(S_L^{-\intercal *} D^{\intercal *} M_S(\zeta S)^\intercal)e_1 + bvec(P_S(\zeta D)^\intercal)e_1 + o(\|D\|_S)
\end{aligned}$$

[30] These block transposes are necessary because the signals are encoded in the columns of the operator S, but each column corresponds to a single shock and multiple signals. Forecasters do not observe individual shocks; they observe individual signals. Forecasting is projecting a single signal onto lags of itself and other signals. So columns of S cannot directly be projected to recover the forecasts. However, columns must be used for projection because the rows of S never contain all of the block entries of S, except in the limit. Therefore the blocks must be transposed so that columns of S correspond to individual signals rather than shocks. These transposes could be avoided by treating causal operators as upper triangular rather than lower triangular (as was the case in earlier versions of the paper) or by having shocks appear on the left-hand side, but this creates more burdensome notation elsewhere.

Separate ζ into the lower triangular (causal) term ζ_C and the strictly upper triangular (noncausal) term ζ_{NC} so that $\zeta = \zeta_C + \zeta_{NC}$:

$$= bvec(M_S D^\intercal S_L^{-\intercal}(\zeta S)^\intercal)e_1 + bvec(S_L^{-\intercal *} D^{\intercal *} M_S (\zeta S)^\intercal)e_1 \\ + bvec(P_S(\zeta_C D)^\intercal)e_1 + bvec(P_S(\zeta_{NC} D)^\intercal)e_1 + o(\|D\|_S)$$

Separating the causal and non-causal components is useful to take advantage of two properties. First, the causal components commute with block transposes, i.e. $(\zeta_C D)^\intercal = D^\intercal \zeta_C^\intercal$. Second, the non-causal component satisfies $bvec(P_S(\zeta_{NC}D)^\intercal)e_1 = bvec(P_S H(P_S D^\intercal)\zeta_{NC}^*)e_1$

$$= bvec(M_S D^\intercal S_L^{-\intercal}(\zeta S)^\intercal)e_1 + bvec(S_L^{-\intercal *} D^{\intercal *} M_S (\zeta S)^\intercal)e_1 \\ + bvec(P_S D^\intercal \zeta_C^\intercal)e_1 + bvec(P_S H(P_S D^\intercal)\zeta_{NC}^*)e_1 + o(\|D\|_S)$$

$$= \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} bvec(M_S D^\intercal)e_1 + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} bvec(D^{\intercal *} M_S[(\zeta S)^\intercal]_+)e_1 \\ + \mathbf{R}^{m_\varepsilon}_{\zeta_C^\intercal} bvec(P_S D^\intercal)e_1 + \mathbf{R}^{m_\varepsilon}_{\zeta_{NC}^\intercal} bvec(P_S H(P_S D^\intercal))e_1 + o(\|D\|_S)$$

$$= \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} bvec(M_S D^\intercal)e_1 + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} bvec(D^{\intercal *} M_S[(\zeta S)^\intercal]_+)e_1 \\ + \mathbf{R}^{m_\varepsilon}_{\zeta_C^\intercal} bvec(P_S D^\intercal)e_1 + \mathbf{R}^{m_\varepsilon}_{\zeta_{NC}^\intercal} bvec(P_S D^\intercal)e_1 + o(\|D\|_S)$$

$$= \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} bvec(M_S D^\intercal)e_1 + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} bvec(D^{\intercal *} M_S[(\zeta S)^\intercal]_+)e_1 + \mathbf{R}^{m_\varepsilon}_{\zeta^\intercal} bvec(P_S D^\intercal)e_1 + o(\|D\|_S)$$

Apply Property 2:

$$= \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} bvec(M_S D^\intercal)e_1 + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} \varrho_{n_\zeta, m_A} bvec(H((M_S[(\zeta S)^\intercal]_+)^\intercal)D^\intercal)e_1 + \mathbf{R}^{m_\varepsilon}_{\zeta^\intercal} bvec(P_S D^\intercal)e_1 + o(\|D\|_S)$$

and because M_S is idempotent:

$$= \mathbf{R}^{m_\varepsilon}_{S_L^{-\intercal}[(\zeta S)^\intercal]_+} bvec(M_S D^\intercal)e_1 + \mathbf{L}^{n_\zeta}_{S_L^{-\intercal *}} \varrho_{n_\zeta, m_A} \mathbf{L}^{m_A}_{H((M_S[(\zeta S)^\intercal]_+)^\intercal)} bvec(M_S D^\intercal)e_1 + \mathbf{R}^{m_\varepsilon}_{\zeta^\intercal} bvec(P_S D^\intercal)e_1 + o(\|D\|_S) \tag{48}$$

Collecting coefficients on $bvec\,(P_S D^\intercal)\,e_1$ and $bvec\,(M_S D^\intercal)\,e_1$ and using the notation from Definition 4 gives

$$bvec\left(\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)^\intercal\right) e_1 = \mathbf{Q}_{P_S} bvec\,(P_S D^\intercal)\,e_1 + \mathbf{Q}_{M_S} bvec\,(M_S D^\intercal)\,e_1 + o(\|D\|_S)$$

Now plug this characterization of $[\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W$ back into Equation (47) and block vectorize:

$$bvec\left(\left(G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right) P_G\right)^\intercal\right) e_1 \\ = \mathbf{L}^{m_A}_{P_G} \mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal} \mathbf{Q}_{P_S} bvec\,(P_S D^\intercal)\,e_1 + \mathbf{L}^{m_A}_{P_G} \mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal} \mathbf{Q}_{M_S} bvec\,(M_S D^\intercal)\,e_1 + o(\|D\|_S) \tag{49}$$

The operators P_S and M_S project onto orthogonally complementary spaces, spanned by – or residual to – the columns of $S^\intercal$, respectively. The vector $bvec\left(D^\intercal\right)e_1$ is the sum of orthogonal components $bvec\left(P_S D^\intercal\right)e_1 = \mathbf{L}_{P_S} bvec\left(D^\intercal\right)$ and $bvec\left(M_S D^\intercal\right)e_1 = \mathbf{L}_{M_S} bvec\left(D^\intercal\right)$. Thus equation (49) becomes

$$\begin{aligned}&bvec\left(\left(G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)P_G\right)^\intercal\right)e_1\\ &\qquad= \left(\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S} + \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}\right) bvec\left(D^\intercal\right)e_1 + o(\|D\|_S)\end{aligned}$$

and taking the limit of $\frac{bvec\left(\left(G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)P_G\right)^\intercal\right)e_1}{\|D\|_S}$ as $\|D\|_S \to 0$ gives the Fréchet derivative. ■

Theorem 8 gives the matrix representation of the Fréchet derivative. Its norm can be computed directly; Appendix C.6 describes how to do so. But first, corollary 2 gives a theoretical property.

Corollary 2 *The norm of the Fréchet derivative satisfies*

$$\|D_{\mathcal{B}}(S)\| = \sqrt{\|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S}\|^2 + \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}\|^2}$$

Proof. By Theorem 8, the norm of the Fréchet derivative is

$$\|D_{\mathcal{B}}(S)\| = \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S} + \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}\|$$

By definition, the operator norm is

$$= \sup_{D \ s.t. \ ||D||_S = 1} \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S} D + \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S} D\|$$

The operators $\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S}$ and $\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}$ act on orthogonal subspaces, spanned by current and past S and orthogonal to it, respectively. Decompose the vector D into these two components:

$$D = D_S + D_{\perp S}$$

The operator norm becomes

$$\|D_{\mathcal{B}}(S)\| = \sup_{D \ s.t. \ ||D_S + D_{\perp S}||_S = 1} \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S} D_S + \mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S} D_{\perp S}\|$$

D_S and $D_{\perp S}$ are orthogonal by construction, so by Lemma 10 the operator norm is

$$\|D_{\mathcal{B}}(S)\| = \sqrt{\|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S}\|^2 + \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}\|^2}$$

■

A.6 Proofs of Characteristics of Stable Equilibria

First, the following Lemma is helpful for proving Theorem 3:

Lemma 5 *If all fixed points of a model contain aggregate signals such that for any fixed point signal vector $\hat{S}$ there is an entry $\hat{S}_i$ satisfying $\hat{S}_i P_G = \hat{S}_i$, then for any $n \geq 1$ the norm of the information feedback operator $G\Theta\Xi(B_{A1}L^{-1} + B_{A0})$ and the component of the Fréchet derivative $\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{R}^{m_\epsilon}_{\zeta^\intercal}\mathbf{L}^{m_A}_{P_S}$ at a fixed point $\hat{S}$ satisfy*

$$\| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \| = \| \left(\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{R}^{m_\epsilon}_{\zeta^\intercal}\mathbf{L}^{m_A}_{P_S}\right)^n \| \tag{50}$$

Proof. For any $n \geq 1$, consider a signal process Y with unit norm satisfying

$$Y = \arg\max_{\|Y\|_s = 1} \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n Y\|_s$$

then we have by definition of the operator norm

$$\| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \| = \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n Y\|_S$$

Consider a single aggregate signal S_G that is generated only by aggregate shocks, i.e. $S_G P_G = S_G$. By assumption, such an S_G is contained in the rows of any fixed point signal $\hat{S}$. Denote the Wold decomposition of the aggregate signal by $S_G = A_G W_G$ where A_G is causally invertible and W_G has variance $\sigma^2_{W_G}$. W_G is white noise, so its autocovariance operator is diagonal: $W_G^* W_G = \sigma^2_{W_G} I$, thus $\sigma^{-1}_{W_G} W_G$ is an isometry. Accordingly, rewriting the operator in terms of the basis $\sigma^{-1}_{W_G} W_G$ does not affect its norm:

$$\begin{aligned}
&= \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n Y \sigma^{-1}_{W_G} W_G\|_S \\
&= \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n Y \sigma^{-1}_{W_G} A_G^{-1} S_G\|_S \\
&= \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \tilde{Y}\|_S
\end{aligned} \tag{51}$$

for the unit norm vector $\tilde{Y} \equiv Y\sigma^{-1}_{W_G} A_G^{-1} S_G$. $\tilde{Y}$ is a $m_A \times m_\varepsilon$ block process that is spanned by current and past signals, and aggregate shocks. Block vectorizing (Appendix A.5.1) preserves the norm:

$$= \|bvec\left(\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \tilde{Y}\right)\|$$

as does block transposing:

$$= \|bvec\left(\left(\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \tilde{Y}\right)^\intercal\right)\|$$

$$= \| \left(\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{R}^{m_\epsilon}_{\zeta^\intercal}\right)^n bvec(\tilde{Y}^\intercal)\|$$

which implies

$$\| \left(G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\right)^n \| = \| \left(\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_S}\right)^n bvec(\tilde{Y}^\intercal)\| \tag{52}$$

per equation (45) and definition 4. By construction, $\tilde{Y}$ is spanned by current and past aggregate signals, so that $bvec(\tilde{Y}^\intercal) = \mathbf{L}^{m_A}_{P_S} bvec(\tilde{Y}^\intercal) = \mathbf{L}^{m_A}_{P_G}\mathbf{L}^{m_A}_{P_S} bvec(\tilde{Y}^\intercal)$. The $\mathbf{L}^{m_A}_{P_G}$ operator commutes with the $\mathbf{R}$ operators in this equation, so:

$$\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_S} bvec(\tilde{Y}^\intercal) = \mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_S}\mathbf{L}^{m_A}_{P_S} bvec(\tilde{Y}^\intercal)$$

Applying this relationship repeatedly, equation (52) becomes:

$$\| \left(G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\right)^n \| = \| \left(\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_S}\mathbf{L}^{m_A}_{P_S}\right)^n bvec(\tilde{Y}^\intercal)\| \tag{53}$$

The initial assumption on Y implied that $\tilde{Y}$ maximizes (51) (subject to $\|\tilde{Y}\|_S = 1$) and thus also (52) and (53). Then by the definition of the operator norm:

$$\| \left(G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\right)^n \| = \| \left(\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{R}^{m_\epsilon}_{\zeta^\intercal}\mathbf{L}^{m_A}_{P_S}\right)^n \|$$

■

Proof of Theorem 3.

Consider the operator norm of the Fréchet derivative:

$$\|\mathcal{D}_\mathcal{B}(\hat{S})\| = \|\mathcal{D}_\mathcal{B}(\hat{S})(P_{\hat{S}} + M_{\hat{S}})\|$$

where $P_{\hat{S}}$ denotes the projection onto the subspace $\mathcal{S}_{P_{\hat{S}}}$ spanned by current and past signals $\hat{S}$, and $M_{\hat{S}} = I - P_{\hat{S}}$ is the projection onto the orthogonal subspace $\mathcal{S}_{M_{\hat{S}}}$.

From the operator norm definition, consider each term:

$$\|\mathcal{D}_\mathcal{B}(\hat{S})P_{\hat{S}}\| = \sup_{y_{\hat{S}}} \frac{\|\mathcal{D}_\mathcal{B}(\hat{S})P_{\hat{S}}y_{\hat{S}}\|_S}{\|y_{\hat{S}}\|_S} \qquad \|\mathcal{D}_\mathcal{B}(\hat{S})M_{\hat{S}}\| = \sup_{y_{\perp\hat{S}}} \frac{\|\mathcal{D}_\mathcal{B}(\hat{S})M_{\hat{S}}y_{\perp\hat{S}}\|_S}{\|y_{\perp\hat{S}}\|_S}$$

The signals $y_{\hat{S}} \in \mathcal{S}_{P_{\hat{S}}}$ and $y_{\perp\hat{S}} \in \mathcal{S}_{M_{\hat{S}}}$ maximizing these norms are orthogonal; therefore, the unit signal $y = y_{\hat{S}} + y_{\perp\hat{S}}$ maximizing $\|\mathcal{D}_\mathcal{B}(\hat{S})y\|_S$ is a linear combination satisfying $\|y_{\hat{S}}\|_S^2 + \|y_{\perp\hat{S}}\|_S^2 = 1$. Maximizing $\|\mathcal{D}_\mathcal{B}(\hat{S})P_{\hat{S}}y_{\hat{S}} + \mathcal{D}_\mathcal{B}(\hat{S})M_{\hat{S}}y_{\perp\hat{S}}\|_S$ subject to this constraint implies the operator norm satisfies

$$\|\mathcal{D}_\mathcal{B}(\hat{S})\| = \sqrt{\|\mathcal{D}_\mathcal{B}(\hat{S})P_{\hat{S}}\|^2 + \|\mathcal{D}_\mathcal{B}(\hat{S})M_{\hat{S}}\|^2} \geq \|\mathcal{D}_\mathcal{B}(\hat{S})P_{\hat{S}}\|$$

Theorem 8 gives the matrix representation for the Fréchet derivative acting on $\mathcal{S}_{P_{\hat{S}}}$ as $\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_{\hat{S}}}\mathbf{L}_{P_{\hat{S}}}$. Therefore

$$\|\mathcal{D}_\mathcal{B}(\hat{S})\| \geq \|\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_{\hat{S}}}\mathbf{L}_{P_{\hat{S}}}\|$$

Apply Lemma 5, letting $n = 1$; the operator norms satisfy

$$\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| = \|\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{R}_{\zeta^\intercal}^{m_\epsilon}\mathbf{L}_{P_S}^{m_A}\|$$

Thus the IFR condition ($\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| < 1$) is necessary for $\|\mathcal{D}_{\mathcal{B}}(\hat{S})\| < 1$ to hold at a fixed point, which by Property 1 is true if and only if the fixed point is signal-stable. ■

Definition 5 *A fixed point $\hat{S}$ is called* locally unique *if there exists a neighborhood $N(S)$ around S such that $\hat{S}$ is the only fixed point in $N(S)$.*

Proof of Theorem 4. The Fréchet derivative is continuous everywhere in $\mathcal{S}_{m_A,m_\varepsilon}$ except at zero, where it is undefined. So if $\|D_{\mathcal{B}}(\hat{S})\| < 1$ then there exists a ball $b(\hat{S})$ around $\hat{S}$ such that $\|D_{\mathcal{B}}(S)\| < 1$ for all $S \in b(\hat{S})$. Therefore $\mathcal{B}$ is a contraction on $b(\hat{S})$ (Kantorovich and Akilov, 1959, p. 661).[31]

$\mathcal{B}$ is also a self-map on $b(\hat{S})$. To see why, consider any $S \in b(\hat{S})$. $\|\mathcal{B}S - \mathcal{B}\hat{S}\|_S < \|S - \hat{S}\|_S$ because $\mathcal{B}$ is a contraction on $b(\hat{S})$. $\hat{S}$ is a fixed point satisfying $\mathcal{B}\hat{S} = \hat{S}$, so $\|\mathcal{B}S - \hat{S}\| < \|S - \hat{S}\|_S$. Therefore $\mathcal{B}S$ is in the ball $b(\hat{S})$.

Finally, $\mathcal{B}$ is a self-map and a contraction on the ball $b(\hat{S})$, therefore the Banach fixed point theorem implies that $\hat{S}$ is the unique fixed point in $b(\hat{S})$. ■

Proof of Corollary 1. A ball $b(\hat{S})$ such that $\|D_{\mathcal{B}}(S)\| < 1\,\forall S \in b(\hat{S})$ exists because $D_{\mathcal{B}}$ as given by Theorem 8 is continuous everywhere that $\|D_{\mathcal{B}}\|$ is finite except 0, and $\|D_{\mathcal{B}}(\hat{S})\| < 1$ because $\hat{S}$ is signal-stable. $\mathcal{B}$ is a contraction mapping on any such ball $b(\hat{S})$ with Lipschitz constant $\max_{S\in b(\hat{S})} \|D_{\mathcal{B}}(S)\| < 1$, therefore by the Banach Contraction Mapping Theorem, $\mathcal{B}^k S_0$ converges to $\hat{S}$. ■

To prove Theorem 5, it is helpful to use the following property of local homeomorphisms (Cartan, 1971, Theorem 4.4.1):

Property 3 *Let $B(a, r)$ be the open ball with radius r around point a in a Banach space E, and let $f : B(a, r) \to E$ be a continuous mapping such that the mapping*

$$\varphi(x) \equiv x - f(x)$$

is a contraction (i.e. it has the k-Lipschitz property for some constant $k < 1$.) Let $f(a) = b$. Then there exists an open set $V \subset B(a, r)$ with $a \in V$ such that f is a homeomorphism of V onto the open ball $B(b, (1 - k)r)$, and the inverse mapping

$$g = f^{-1} : B(b, (1 - k)r) \to B(a, r)$$

has the $\frac{1}{1-k}$-Lipschitz property.

[31] This property is reported in English in Holtzman (1968).

Proof of Theorem 5. $\hat{S} \neq 0$ is signal-stable, so $1 > \|D_{\mathcal{B}}(\hat{S})\|$. As in the Proof of Theorem 4, continuity of the norm $\|D_{\mathcal{B}}\|$ away from zero implies there exists some ball $b(\hat{S}, r^*)$ with radius r^* such that $\mathcal{B}$ is a contraction on $b(\hat{S}, r^*)$ for any Lipschitz coefficient $k_{\mathcal{B}} \in (\|D_{\mathcal{B}}(\hat{S})\|, 1)$, and $\|D_{\mathcal{B}}(S)\| < 1$ for all $S \in b(\hat{S}, r^*)$. The truncation operation in $\mathcal{B}_\tau$ implies $\|D_{\mathcal{B}}(S)\| \geq \|D_{\mathcal{B}_\tau}(S)\|$ for all non-zero $S \in \mathcal{S}_{m_A, m_\varepsilon}$, so $\mathcal{B}_\tau$ is also a contraction on any ball $b(\hat{S}, r)$ with $r \leq r^*$.

Per Lemma 3, for any r there exists a K such that for any $\tau > K$

$$\|\mathcal{B}_\tau(\hat{S}) - \mathcal{B}(\hat{S})\|_{\mathcal{S}} < (1 - k_{\mathcal{B}})r$$

Define $\mathcal{C}_\tau \equiv I - \mathcal{B}_\tau$ and $\mathcal{C} \equiv I - \mathcal{B}$. It must also be the case that:

$$\|\mathcal{C}_\tau(\hat{S}) - \mathcal{C}(\hat{S})\|_{\mathcal{S}} < (1 - k_{\mathcal{B}})r$$

Property 3 implies that if $r \leq r^*$, there exists a homeomorphism $g_\tau : B_\tau(\mathcal{C}_\tau(\hat{S}), (1 - k_{\mathcal{B}})r) \to B_\tau(\hat{S}, r)$. $\hat{S}$ is a fixed point of $\mathcal{B}$, so $\mathcal{C}(\hat{S}) = 0$. Thus $0 \in B_\tau(\mathcal{C}_\tau(\hat{S}), (1 - k_{\mathcal{B}})r)$ and $g_\tau(0) \in B_\tau(\hat{S}, r)$, so there exists a fixed point $\hat{S}_\tau = \mathcal{B}_\tau(\hat{S}_\tau)$ such that $\|\hat{S}_\tau - \hat{S}\|_{\mathcal{S}} < r$. $\hat{S}_\tau$ must be signal-stable because $\hat{S}_\tau \in b(\hat{S}, r)$ implies $\|D_{\mathcal{B}_\tau}(\hat{S})\| < 1$. This proves there exists a sequence of signal-stable fixed points $\hat{S}_\tau$ such that $\lim_{\tau \to \infty} \hat{S}_\tau = \hat{S}$. ■

A.7 Stable Uniqueness

Proving Theorem 6 requires some notation and intermediate results.

First, I define a set which includes all signal-stable equilibrium fixed points. Let $\mathcal{Y}$ denote the bounded set of $m_A \times m_\varepsilon$-block signals with S-norm $< R_S$ around which $I - \mathcal{B}$ is signal-stable:

$$\mathcal{Y} \equiv \{S \in \mathcal{S}_{m_\varepsilon, m_A} : \|S\|_S < R_S, \ \|D_{\mathcal{B}}(S)\| < 1\}$$

where $R_S = \frac{\|S_X\|_S}{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|}$.

Additionally, let $\mathcal{Y}_\tau$ denote the subset of $\mathcal{Y}$ in the image of P_τ, i.e. the subset of signals truncated at order τ or less.

Lemma 6 *If a signal with Wold representation $S = AW$ has a matrix (i.e. block diagonal) forecast error operator W, and if Information Feedback Regularity holds, then $\|D_{\mathcal{B}}(S)\| < 1$*

Proof. The operator $\mathcal{B}(S)$ is given by

$$\mathcal{B}(S) = S_X + G\Theta[\Xi(B_{A1}L^{-1} + B_{A0})A]_+ WP_G$$

If W is a matrix (i.e. block diagonal operator) then it commutes with the annihilation operator:

$$= S_X + G\Theta[\Xi(B_{A1}L^{-1} + B_{A0})AWP_G]_+$$

$$= S_X + G\Theta[\Xi(B_{A1}L^{-1} + B_{A0})SP_G]_+$$

which is a linear operator $\overline{\mathcal{B}}$ on S with norm $\|\overline{\mathcal{B}}\| = \|\mathcal{G}\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|$. If Information Feedback Regularity holds, then this linear operator has norm $\|\overline{\mathcal{B}}\| < 1$ thus

$$\|D_{\mathcal{B}}(S)\| = \|\overline{\mathcal{B}}\| < 1$$

■

In the next lemma, $m_{\varepsilon,I}$ denotes the number of idiosyncratic shock dimensions.

Lemma 7 *If Information Feedback Regularity holds, and $m_A < m_{\varepsilon,I}$ then $\mathcal{Y}_\tau \subset \mathcal{S}_{m_A,m_\varepsilon}$ is a path-connected Banach manifold.*

Proof. To show path-connectedness, the proof constructs a series of homotopies that connect any two signals in $\mathcal{Y}_\tau$

The first step of the proof is to define a homotopy through $\mathcal{Y}_\tau$ that begins with a signal $S^1 = A^1W^1$ and ends with the signal $S^2 = A^1W^2$, by transforming the white noise process W^1 into a second process W^2 with the same variance Σ_W but determined entirely by idiosyncratic shocks.

The norm of the Fréchet derivative of $\mathcal{B}_\tau$ around S^1 is

$$\|\mathcal{D}_{\mathcal{B}}(S^1)\| = \lim_{\Delta\to 0} \sup_{D^1 \text{ given } \|D^1\|_S=\Delta} \|\mathcal{B}(S^1 + D^1) - \mathcal{B}(S^1)\|_{\mathcal{S}}$$

Per Theorem 8, a norm-maximizing deviation D^1 is the linear combination of a component spanned by W^1 and a component orthogonal to W^1. Let A^DW^1 denote the first component, and let $A^D_\perp W^1_\perp$ denote the latter, with $W^1_\perp$ the Wold representation's white noise basis with variance $\Sigma_\perp$.

With this decomposition for the deviation $D^1 = A^DW^1 + A^D_\perp W^1_\perp$, the square of the norm (A.7) becomes

$$\begin{aligned}
&\lim_{\Delta\to 0} \sup_{D^1 \text{ given } \|D^1\|_S=\Delta} \|\mathcal{B}(S^1 + D^1) - \mathcal{B}(S^1)\|^2_{\mathcal{S}} = \\
&\lim_{\Delta\to 0} \|G\Theta\left([\Xi[(B_{A1}L^{-1} + B_{A0})A^D]_+]_+W^1P_G + [\Xi[(B_{A1}L^{-1} + B_{A0})A^D_\perp]_+]_+W^1_\perp P_G\right. \\
&\qquad\qquad \left. - [\Xi[(B_{A1}L^{-1} + B_{A0})A]_+]_+W^1P_G\right)\|^2_{\mathcal{S}}
\end{aligned}$$

$$\begin{aligned}
= &\lim_{\Delta\to 0} \sup_{D^1 \text{ s.t. } \|D^1\|_S=\Delta} \|G\Theta\left([\Xi[(B_{A1}L^{-1} + B_{A0})(A^D - A)]_+]_+W^1P_G\right. \\
&\qquad\qquad \left. + [\Xi[(B_{A1}L^{-1} + B_{A0})A^D_\perp]_+]_+W^1_\perp P_G\right)\|^2_{\mathcal{S}}
\end{aligned}$$

There are many possible orthogonal white noise processes that could serve as $W^1_\perp$. Let $W^1_{\perp,G}$ and $W^1_{\perp,I}$ denote such processes that are entirely aggregate and idiosyncratic respectively. It is always possible to find such processes that are white noise and orthogonal to W^1; for example, take the white noise from the Wold representations of

$[L^{-1}W^1P_G]_+$ and $[L^{-1}W^1(I-P_G)]_+$ respectively to construct such candidates. The norm-maximizing $W^1_\perp$ must be entirely aggregate, i.e. it must be in the image of the projection P_G to satisfy $W^1_{\perp,G}P_G = W^1_{\perp,G}$. By construction $W^1_{\perp,G}$ is orthogonal to W^1 and thus W^1P_G, so the norm equality continues:

$$= \lim_{\Delta\to 0} \sup_{D^1 \text{ s.t. } \|D^1\|_S=\Delta} \|G\Theta[\Xi[(B_{A1}L^{-1}+B_{A0})(A^D-A)]_+]_+ W^1P_G\|^2_{\mathcal{S}} + \\ \|G\Theta[\Xi[(B_{A1}L^{-1}+B_{A0})A^D_\perp]_+]_+ W^1_{\perp,G}\|^2_{\mathcal{S}}$$

W^1P_G is not necessarily white noise; define its Wold representation as $W^1P_G = A^1_{WG}W^1_{WG}$. For $s\in[0,1]$ define the homotopy $S^1_s = (1-s)A^1W^1 + sA^1(W^1(I-P_G) + A^1_{WG}W^1_{\perp,I})$ for some $W^1_{\perp,I}$ with the same variance as W^1_{WG}. S^1_s has the same autocovariance function as S^1 and is orthogonal to $W^1_{\perp,G}$ for all s. But its square of the norm (A.7) is now given by

$$\|\mathcal{D}_{\mathcal{B}}(S^1_s)\|^2 = \lim_{\Delta\to 0} \sup_{D^1_s \text{ s.t. } \|D^1_s\|_S=\Delta} \|G\Theta[\Xi[(B_{A1}L^{-1}+B_{A0})(A^D-A)]_+]_+(1-s)W^1P_G\|^2_{\mathcal{S}} + \\ \|G\Theta[\Xi[(B_{A1}L^{-1}+B_{A0})A^D_\perp]_+]_+ W_{\perp,G}\|^2_{\mathcal{S}} \quad (54)$$

where $W_{\perp,G}$ denotes a norm-maximizing aggregate orthogonal deviation at each point in the homotopy (A^D and $A^D_\perp$ are unchanged).

Define $S^2 = S^1_1$, i.e. the end point of the homotopy. This signal has the wold representation $S^2 = A^1W^2$ and W^2 is entirely idiosyncratic i.e. $W^2P_G = 0$. The norm (54) is decreasing as $s\to 1$: for all s, $\|\mathcal{D}_{\mathcal{B}}(S^1_s)\| \le \|\mathcal{D}_{\mathcal{B}}(S^1)\|$. Therefore if $S^1\in\mathcal{Y}$, then it is path-connected to S^2 through $\mathcal{Y}$. Moreover, $S^1 = A^1W^1 \in \mathrm{im}(P_\tau)$, so $S^1(I-P_G) = A^1W^1(I-P_G)\in\mathrm{im}(P_\tau)$ and $S^1P_G = A^1A^1_{WG}W^1_{WG}\in\mathrm{im}(P_\tau)$. Therefore by choosing $W^1_{\perp,I}$ such that $A^1A^1_{WG}W^1_{\perp,I}\in\mathrm{im}(P_\tau)$, the homotopy $S^1_s\in\mathrm{im}(P_\tau)$ for all s. Thus if $S^1\in\mathcal{Y}_\tau$, then it is path-connected to S^2 through $\mathcal{Y}_\tau$.

The second step of the proof is to show that the signal $S^2 = A^1W^2$ is path-connected through $\mathcal{Y}_\tau$ to the signal $S^3 = A^1W^3$, where W^3 is a block diagonal white noise process with the same variance $\Sigma_W = C_WC_W^*$. At S^2, the square of the operator norm is simple, because W^2 is entirely idiosyncratic so the first term from equation (54) drops out:

$$\|\mathcal{D}_{\mathcal{B}}(S^2)\|^2 = \lim_{\Delta\to 0} \sup_{D^2 \text{ s.t. } \|D^2\|_S=\Delta} \|G\Theta[\Xi[(B_{A1}L^{-1}+B_{A0})A^D_\perp]_+]_+ W_{\perp,G}\|^2_{\mathcal{S}} \quad (55)$$

and now the deviation must be given by $D^2 = A^D_\perp W_{\perp,G}$. Because S^2 is entirely idiosyncratic, the right-hand side norm is achieved by any choice of $W_{\perp,G}$, appropriately scaled. Thus, this is the Fréchet derivative norm for any $S^i = A^1W^i$ so long as W^i is idiosyncratic with variance Σ_W. $C_W^{-1}W^i$ has variance matrix I, so $C_W^{-1}W^i$ is a co-isometry, and $C_W^{-1}W^i$ is rational because τ is finite. $C_W^{-1}W^i$ only has non-zero terms in the $m_{\varepsilon,I} > m_A$ idiosyncratic dimensions. Within the idiosyncratic dimensions alone, the adjoint of this operator is a "tall" rational isometry with $m_{\varepsilon,I}\times m_A$ blocks; Jury

(2023) proves this set of isometries is path-connected. Also in this set is $C_W^{-1} W_3$, where W_3 is a $m_A \times m_\varepsilon$ matrix-valued white noise process (i.e. its block Toeplitz operator is block diagonal) with variance matrix Σ_W and is also entirely idiosyncratic. Therefore $S^2 = A^1 W^2$ is path-connected through $\mathcal{Y}_\tau$ to the signal $S^3 = A^1 W^3$

For the third step of the proof, let W_4 denote a $m_A \times m_\varepsilon$ matrix-valued white noise process (i.e. its block Toeplitz operator is block diagonal) with variance matrix I. Lemma 6 implies that all signals whose Wold representation features a matrix-valued white noise operator are in $\mathcal{Y}$. And if S^1 is of order τ or less, then A^1 is of order τ or less. Therefore $S^3 = A^1 W^3$ is path-connected to $S^4 = A^1 W^4$ through $\mathcal{Y}_\tau$ by the homotopy $A^1(sW^4 + (1-s)W^3)$ for $s \in [0,1]$, and S^4 is path-connected to W^4 through $\mathcal{Y}_\tau$ by the homotopy $(sI + (1-s)A^1)W^4$ $s \in [0,1]$. Therefore, all signals in $\mathcal{Y}_\tau$ are path-connected to W^4 and thus each other.

■

The proof strategy for Theorem 6 requires a model to have more idiosyncratic shocks than signals, so in cases where this is not satisfied, modify the model's shock space to have additional idiosyncratic "sunspot" dimensions. These sunspot shocks do not affect the exogenous signal process S_X, and because they are idiosyncratic (i.e. the space spanned by the additional dimensions is in the kernel of P_G) they cannot affect the endogenous signals either. Therefore this modification introduces no new fixed points, and has no effect on the norm of the Fréchet derivative or signal-stability, but is useful so that Lemma 7 applies. With this addition, the shock dimensions are now of size m_ε^*. In the proof below, it is assumed that the operators (e.g. $\mathcal{B}_\tau$) and subspaces (e.g. $\mathcal{Y}_\tau$) are defined on this modified space $\mathcal{S}_{m_A, m_\varepsilon^*}$

Proof of Theorem 6. By construction, Lemma 1 implies that all signal-stable fixed points are in $\mathcal{Y}$.

$\mathcal{Y}_\tau$ is finite-dimensional and bounded, so $I - \mathcal{B}_\tau : \mathcal{Y}_\tau \to \mathcal{S}_{m_A, m_\varepsilon^*}$ is proper.

$\mathcal{B}_\tau = \mathcal{B} P_\tau$ where P_τ is a projection operator, so $\|D_{\mathcal{B}_\tau}(S)\| \leq \|D_{\mathcal{B}}(S)\|$. Therefore if $S \in \mathcal{Y}$, then $\|D_{\mathcal{B}_\tau}(S)\| < 1$ and by the inverse function theorem, $I - \mathcal{B}_\tau$ is a local homeomorphism on $\mathcal{Y}$ and thus on $\mathcal{Y}_\tau$.

$I - \mathcal{B}_\tau : \mathcal{Y}_\tau \to \mathcal{S}_{m_A, m_\varepsilon^*}$ is a proper local homeomorphism and $\mathcal{S}_{m_A, m_\varepsilon^*}$ is connected, so by the Browder Theorem (Browder, 1954) $I - \mathcal{B}_\tau$ is a covering projection with finite fiber.[32] If Information Feedback Regularity holds, then $\mathcal{Y}_\tau$ is path-connected by Lemma 7 and $\mathcal{S}_{m_A, m_\varepsilon^*}$ is simply connected because it is a Banach space. Therefore by a standard monodromy theorem, $I - \mathcal{B}_\tau : \mathcal{Y}_\tau \to \mathcal{S}_{m_A, m_\varepsilon^*}$ is a global homeomorphism (Katriel, 1994, Thm 4.1).

There is at most one $S \in \mathcal{Y}_\tau$ such that $S = \mathcal{B}_\tau(S)$, and $\mathcal{Y}_\tau$ contains all signal-stable fixed points of $\mathcal{B}_\tau$, so there is at most one signal-stable approximate fixed point of order τ. It remains to be proven that there is at most one signal-stable fixed point of the untruncated operator $\mathcal{B}$.

Suppose towards a contradiction that there are multiple signal-stable fixed points $\hat{S}_i$ of $\mathcal{B}$, indexed by i. By Theorem 4, these points must be locally unique. By Theorem

[32]For unfamiliar economists, Gutú (2017) provides an accessible summary of these properties.

5, each of these points has a convergent sequence of signal-stable approximate fixed points $\hat{S}_{i,\tau}$, indexed by the truncation order τ. Select a scalar r such that there are disjoint balls of radius r around each fixed point. Select an approximation order τ^*, such that each ball $b(\hat{S}_i, r)$ contains an element of the sequence $\hat{S}_{i,\tau^*}$. The ball disjointedness implies that the operator $\mathcal{B}_{\tau^*}$ has multiple signal-stable approximate fixed points. This is a contradiction; therefore there can be at most one signal-stable fixed point of $\mathcal{B}$. ■

A.8 Proofs Related to the Sufficient Condition

Proof of Theorem 7. Lemma 8 says that if the Sufficient Idiosyncrasy condition (SIC) holds then all fixed points must be signal-stable. SIC also implies IFR:

$$\mathrm{r}(\Sigma_I^{-1}) < \frac{1 - \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2}{4R_N^2}$$

$$\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + 4R_N^2\mathrm{r}(\Sigma_I^{-1}) < 1$$

$$\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 < 1 - 4R_N^2\mathrm{r}(\Sigma_I^{-1}) < 1$$

therefore $\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| < 1$, which means that Theorem 6 holds, so there is at most one signal-stable fixed point. ■

The proof of Theorem 7 is only straightforward because of Lemma 8.

Lemma 8 *If the Sufficient Idiosyncrasy Condition holds, then all fixed points satisfying $\hat{S} = \mathcal{B}(\hat{S})$ must be signal stable.*

Proof. Consider a signal perturbation $D \equiv \hat{S}^\Delta - \hat{S}$. Decompose the signal deviation $D = D_{\hat{S}} + D_{\perp\hat{S}}$ into the component $D_{\hat{S}}$ that is spanned by lags of $\hat{S}$, and the orthogonal component $D_{\perp\hat{S}}$. As in equation (47), the perturbed difference $\mathcal{B}(\hat{S}^\Delta) - \mathcal{B}(\hat{S})$ is given by

$$\mathcal{B}(\hat{S}^\Delta) - \mathcal{B}(\hat{S}) = G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right) P_G \tag{56}$$

Take the signal norm:

$$\|\mathcal{B}(\hat{S}^\Delta) - \mathcal{B}(\hat{S})\|_S = \|G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right) P_G\|_S$$

P_G is a projection operator, so this norm is bounded above by

$$\leq \|G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)\|_S = \|G\Theta\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)\bar{e}_{m_\varepsilon}\|_S$$

where the operator $\bar{e}_m$ denotes the block vector consisting of the first m basis vectors, e.g. $\bar{e}_{m_\varepsilon} = \left(e_1, ..., e_{m_\varepsilon}\right)$. The equation then follows directly from the definition of the signal norm, which only depends on the first block column of an operator.

$[\zeta A]_+ W$ is the projection of the noncausal signals $\zeta\hat{S}$ onto current and past $\hat{S}$. To write this projection step as a linear operator, it is necessary to transpose the blocks of the signals. $\hat{S}^\intercal$ is still lower triangular block Toeplitz, but each block is transposed. This approach is useful because Lemma 4 can simplify the block-transposed representation. As in Appendix A.5.2, $P_{\hat{S}} = \hat{S}^\intercal(\hat{S}^{\intercal *}\hat{S}^\intercal)^{-1}\hat{S}^{\intercal *}$ denotes the projection onto the columns of $\hat{S}^\intercal$, $M_{\hat{S}} \equiv I - P_{\hat{S}}$ denotes the residual projection and $\hat{S}_L^{-\intercal} \equiv (\hat{S}^{\intercal *}\hat{S}^\intercal)^{-1}\hat{S}^{\intercal *}$ denotes the left inverse of $\hat{S}^\intercal$.

Block-transpose the $\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)\bar{e}_{m_\varepsilon}$ block vector to continue the equality:

$$= \|G\Theta\left(\left([\zeta A^\Delta]_+ W^\Delta - [\zeta A]_+ W\right)^\intercal \bar{e}_{m_A}\right)^\intercal\|_S$$

$$= \|G\Theta\left(D^\intercal_{\perp\hat{S}}\hat{S}_L^{-\intercal}[(\zeta\hat{S})^\intercal]_+\bar{e}_{m_A} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}[(\zeta\hat{S})^\intercal]_+\bar{e}_{m_A} + P_{\hat{S}}[(\zeta D_{\hat{S}})^\intercal]_+\bar{e}_{m_A}\right)^\intercal\|_S + o(\|D\|_S)$$

by Lemma 4, using the $D = D_{\hat{S}} + D_{\perp\hat{S}}$ decomposition, where $D^\intercal_{\hat{S}} = P_{\hat{S}}D^\intercal$ and $D_{\perp\hat{S}} = M_{\hat{S}}D^\intercal$. Then apply the triangle inequality to bound $\mathcal{B}(\hat{S}^\Delta) - \mathcal{B}(\hat{S})$:

$$\begin{aligned}&\|\mathcal{B}(\hat{S}^\Delta) - \mathcal{B}(\hat{S})\|_S \leq \\ &\|G\Theta\left(D^\intercal_{\perp\hat{S}}\hat{S}_L^{-\intercal}[(\zeta\hat{S})^\intercal]_+ + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}[(\zeta\hat{S})^\intercal]_+\bar{e}_{m_A}\right)^\intercal\|_S + \|\left(P_{\hat{S}}[(\zeta D_{\hat{S}})^\intercal]_+\bar{e}_{m_A}\right)^\intercal\|_S + o(\|D\|_S)\end{aligned} \tag{57}$$

First consider the $D_{\hat{S}}$ term in inequality (57), which by the signal norm definition is

$$\|G\Theta\left(P_{\hat{S}}[(\zeta D_{\hat{S}})^\intercal]_+\bar{e}_{m_A}\right)^\intercal\|_S = \|G\Theta\left(P_{\hat{S}}[(\zeta D_{\hat{S}})^\intercal]_+\right)^\intercal\|_S$$

then evaluate the norm-maximizing deviation:

$$\leq \sup_{d\, s.t. \|d\|_S = \|D_{\hat{S}}\|_S} \|G\Theta\left(P_{\hat{S}}[(\zeta d)^\intercal]_+\right)^\intercal\|_{\hat{S}}$$

$$\leq \sup_{d\, s.t. \|d\|_S = \|D_{\hat{S}}\|_S} \|G\Theta\left(P_{\hat{S}}(\zeta d)^\intercal\right)^\intercal\|_S \leq \sup_{d\, s.t. \|d\|_S = \|D_{\hat{S}}\|_S} \|G\Theta\left((\zeta d)^\intercal\right)^\intercal\|_{\hat{S}}$$

because to be norm-maximizing, applying the annihilator must not be norm-reducing (the norm-maximizing d is a signal lagged arbitrarily far into the past), and then applying $P_{\hat{S}}$ must not be norm-reducing (the norm-maximizing d is spanned by current and past $\hat{S}$). Continuing the inequality:

$$= \sup_{d\, s.t. \|d\|_S = \|D_{\hat{S}}\|_S} \|G\Theta\zeta d\|_S = \|G\Theta\zeta\|\|D_{\hat{S}}\|_S$$

by definition of the operator norm. To summarize this chain of inequalities, the $D_{\hat{S}}$ component of inequality (57) is bounded by

$$\|G\Theta\left(P_{\hat{S}}[(\zeta D_{\hat{S}})^\intercal]_+\bar{e}_m\right)^\intercal\|_S \leq \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|\|D_{\hat{S}}\|_S \tag{58}$$

which uses the definition $\zeta = \Xi(B_{A1}L^{-1} + B_{A0})$.

Next consider the $D_{\perp\hat{S}}$ terms from inequality (57):

$$\|G\Theta\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal}[(\zeta\hat{S})^{\intercal}]_+\bar{e}_m + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}[(\zeta\hat{S})^{\intercal}]_+\bar{e}_m\right)^{\intercal}\|_S$$
$$= \|G\Theta\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal}[(\zeta\hat{S})^{\intercal}]_+ + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}[(\zeta\hat{S})^{\intercal}]_+\right)^{\intercal}\|_S$$
$$= \|G\Theta\left(\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)[(\zeta\hat{S})^{\intercal}]_+\right)^{\intercal}\|_S$$

Lemma 9 implies

$$\|G\Theta\left(\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)[(\zeta\hat{S})^{\intercal}]_+\right)^{\intercal}\|_S$$
$$\leq \|\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)\|\left(\|G\Theta\zeta\|\|\hat{S}P_G\|_S + \vartheta_I\right)$$

$\hat{S}$ is a fixed point and IFR holds, so by Lemma 2, its aggregate component is bounded by $\|\hat{S}P_G\|_S \leq \frac{\|S_X P_G\|_S + \vartheta_I}{1-\|G\Theta\zeta\|}$. Therefore the inequality becomes

$$\|G\Theta\left(\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)[(\zeta\hat{S})^{\intercal}]_+\right)^{\intercal}\|_S$$
$$\leq \|\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)\|\left(\|G\Theta\zeta\|\frac{\|S_X P_G\|_S + \vartheta_I}{1-\|G\Theta\zeta\|} + \vartheta_I\right)$$
$$= \|\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)\|R_N$$

The operator norm $\|D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\|$ is relatively easy to characterize because the operator is self-adjoint, so its operator norm is given by its spectral radius:

$$\|D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\| = \mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal} + \hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right)$$
$$\leq \mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal}\right) + \mathrm{r}\left(\hat{S}_L^{-\intercal *}D^{\intercal *}_{\perp\hat{S}}\right) = 2\mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal}\right)$$

because the spectral radius is sub-additive and adjoint-invariant.[33] Next, consider the block-transpose of the Wold representation $\hat{S}^{\intercal} = W^{\intercal}A^{\intercal}$, and recognize that $W^{\intercal}C_W^{-*}$ is an isometry because $W^{\intercal *}W^{\intercal} = \Sigma_W = C_W C_W^*$. Therefore the definition $\hat{S}_L^{-\intercal} = (\hat{S}^{\intercal *}\hat{S}^{\intercal})^{-1}\hat{S}^{\intercal *}$ implies $\hat{S}_L^{-\intercal} = (A^{\intercal *}W^{\intercal *}W^{\intercal}A^{\intercal})^{-1}A^{\intercal *}W^{\intercal *} = A^{-\intercal}C_W^{-*}C_W^{-1}A^{-\intercal *}A^{\intercal *}W^{\intercal *}$ which simplifies to $A^{-\intercal}C_W^{-*}(W^{\intercal}C_W^{-*})^*$. Thus:

$$2\mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}\hat{S}_L^{-\intercal}\right) = 2\mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}A^{-\intercal}C_W^{-*}(W^{\intercal}C_W^{-*})^*\right) \leq 2\mathrm{r}\left(D^{\intercal}_{\perp\hat{S}}A^{-\intercal}C_W^{-*}\right)$$

[33] This relationship actually holds with equality, although that is unnecessary to show for this proof.

because $W^{\intercal} C_W^{-*}$ is an isometry. $D_{\perp \hat{S}}^{\intercal}$ and $A^{-\intercal} C_W^{-*}$ are both block lower triangular, so the block transpose can be reversed, i.e.

$$2\mathrm{r}\left(D_{\perp \hat{S}}^{\intercal} A^{-\intercal} C_W^{-*}\right) = 2\mathrm{r}\left(C_W^{-1} A^{-1} D_{\perp \hat{S}}\right)$$

The diagonal blocks of the operator $C_W^{-1} A^{-1} D_{\perp \hat{S}}$ are $C_W^{-1} D_{\perp \hat{S},0}$, because the diagonal blocks of A are identities. The spectral radius of a block triangular operator is the largest spectral radius of the main diagonal blocks, therefore

$$2\mathrm{r}\left((AC_W)^{-1} D_{\perp \hat{S}}\right) = 2\mathrm{r}\left(C_W^{-1} D_{\perp \hat{S},0}\right)$$

where $D_{\perp \hat{S},0}$ is the main diagonal block of $D_{\perp \hat{S}}$. And the spectral radius is bounded above by the norm:

$$2\mathrm{r}\left(C_W^{-1} D_{\perp \hat{S},0}\right) \leq 2\|C_W^{-1} D_{\perp \hat{S},0}\| \leq 2\|C_W^{-1}\|\|D_{\perp \hat{S},0}\|$$

To summarize, this chain of inequalities implies the bound

$$\|G\Theta\left(D_{\perp \hat{S}}^{\intercal} \hat{S}_L^{-\intercal}[(\zeta \hat{S})^{\intercal}]_+ \bar{e}_m + \hat{S}_L^{-\intercal *} D_{\perp \hat{S}}^{\intercal *}[(\zeta \hat{S})^{\intercal}]_+ \bar{e}_m\right)^{\intercal}\|_S \leq 2R_N \|C_W^{-1}\|\|D_{\perp \hat{S},0}\| \tag{59}$$

Next I will bound each of the right-hand side terms.

To characterize C_W^{-1}, first decompose the innovation process as $W = W_I + W_{\perp I}$ where $W_I = S_{X,0}(I - P_G)$, i.e. the component of the exogenous process S_X that is due to contemporaneous idiosyncratic shocks. $W_{\perp I}$ is the residual. Because these components are orthogonal, the forecast error variance satisfies

$$\Sigma_W = \Sigma_I + \Sigma_{\perp I} \tag{60}$$

where $\Sigma_I = W_I W_I^*$ and $\Sigma_{\perp I} = W_{\perp I} W_{\perp I}^*$. All variance matrices in equation (60) are real, symmetric, and positive-semi definite. Therefore, the minimum eigenvalues $(\underline{\lambda}_W, \underline{\lambda}_I, \underline{\lambda}_{\perp I})$ of the respective matrices $(\Sigma_W, \Sigma_I, \Sigma_{\perp I})$ satisfy

$$0 < \underline{\lambda}_I + \underline{\lambda}_{\perp I} \leq \underline{\lambda}_W$$

The SIC implies that the variance matrix Σ_I is positive definite, so $0 < \underline{\lambda}_I$ Invert:

$$0 < \frac{1}{\underline{\lambda}_W} \leq \frac{1}{\underline{\lambda}_I + \underline{\lambda}_{\perp I}} \leq \frac{1}{\underline{\lambda}_I}$$

The maximum eigenvalues of the inverse matrices are the inverse minimum eigenvalues, which immediately implies that the spectral radii satisfy

$$\sqrt{\mathrm{r}\left(\Sigma_W^{-1}\right)} \leq \sqrt{\mathrm{r}\left(\Sigma_I^{-1}\right)}$$

and because $C_W^{-*}C_W^{-1} = \Sigma_W^{-1}$ and $C_I^{-*}C_I^{-1} = \Sigma_I^{-1}$, these square roots are equivalent to the matrix norms:

$$\|C_W^{-1}\| \leq \|C_I^{-1}\| \tag{61}$$

To characterize $\|D_{\perp\hat{S},0}\|$, use that the operator norm of a matrix is bounded above by its Frobenius norm:

$$\|D_{\perp\hat{S},0}\| \leq \|D_{\perp\hat{S},0}\|_F \leq \|D_{\perp\hat{S}}\|_S \tag{62}$$

where $\|\cdot\|_F$ denotes the Frobenius norm, i.e. the root sum of squares of all entries in $D_{\perp\hat{S},0}$. This must be less than or equal to $\|D_{\perp\hat{S}}\|_S$, the root sum of squares of all entries in all $\{D_{\perp\hat{S},j}\}_{j=0}^{\infty}$ matrices (see equation (40)).

By applying inequalities (61) and (62), inequality (59) becomes

$$\|G\Theta\left(D_{\perp\hat{S}}^{\intercal}\hat{S}_L^{-\intercal}[(\zeta\hat{S})^{\intercal}]_+\bar{e}_m + \hat{S}_L^{-\intercal*}D_{\perp\hat{S}}^{\intercal*}[(\zeta\hat{S})^{\intercal}]_+\bar{e}_m\right)^{\intercal}\|_S \leq 2R_N\|C_I^{-1}\|\|D_{\perp\hat{S}}\|_S$$

Combining this result with inequality (58), inequality (57) becomes

$$\|\mathcal{B}(\hat{S}^{\Delta}) - \mathcal{B}(\hat{S})\|_S \leq \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|\|D_{\hat{S}}\|_S + 2R_N\|C_I^{-1}\|\|D_{\perp\hat{S}}\|_S + o(\|D\|_S) \tag{63}$$

The deviations $D_{\hat{S}}$ and $D_{\perp\hat{S}}$ are orthogonal, i.e. $\|D_{\hat{S}}\|_S^2 + \|D_{\perp\hat{S}}\|_S^2 = \|D\|_S^2$. By Lemma 10, the linear combination of deviations maximizing the right-hand side of inequality (63) gives a maximum value $\sqrt{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + \left(2R_N\|C_I^{-1}\|\right)^2}\|D\|_S$. Therefore the inequality becomes

$$\|\mathcal{B}(\hat{S}^{\Delta}) - \mathcal{B}(\hat{S})\|_S < \sqrt{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + \left(2R_N\|C_I^{-1}\|\right)^2}\|D\|_S + o(\|D\|_S)$$

and taking the limit as $\|D\|_S \to 0$ bounds the Fréchet derivative:

$$\begin{aligned}\|\mathcal{D}_{\mathcal{B}(\hat{S})}\| &< \sqrt{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + \left(2R_N\|C_I^{-1}\|\right)^2} \\ &= \sqrt{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + 4R_N^2\mathrm{r}(\Sigma_I^{-1})}\end{aligned}$$

again using that $\|C_I^{-1}\| = \sqrt{\mathrm{r}(\Sigma_I^{-1})}$. The Sufficient Idiosyncrasy Condition implies $\sqrt{\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|^2 + 4R_N^2\mathrm{r}(\Sigma_I^{-1})} < 1$, and thus $\|\mathcal{D}_{\mathcal{B}(\hat{S})}\| < 1$. So if $\hat{S}$ is a fixed point, it must be stable. ■

Lemma 9 *If the signal S is a fixed point, then for any operator Φ*

$$\|G\Theta\left(\Phi[(\zeta S)^{\intercal}]_+\right)^{\intercal}\|_S \leq \|\Phi\|\left(\|G\Theta\zeta\|\|SP_G\|_S + \vartheta_I\right)$$

where ϑ_I is defined as in Lemma 2.

Proof. Decompose the forecast $[(\zeta S)^\intercal]_+$ into components driven by aggregate and idiosyncratic shocks: $[(\zeta S)^\intercal]_+ = [(\zeta S P_G)^\intercal]_+ + [(\zeta S(I - P_G))^\intercal]_+$. By the triangle inequality:

$$\begin{aligned}
&\|G\Theta\left(\Phi[(\zeta S)^\intercal]_+\right)^\intercal\|_S \\
&\qquad \leq \|G\Theta\left(\Phi[(\zeta S P_G)^\intercal]_+\right)^\intercal\|_S + \|G\Theta\left(\Phi[(\zeta S(I-P_G))^\intercal]_+\right)^\intercal\|_S
\end{aligned} \tag{64}$$

First, consider the aggregate term:

$$\begin{aligned}
\|G\Theta\left(\Phi[(\zeta S P_G)^\intercal]_+\right)^\intercal\|_S &\leq \sup_{s\, s.t. \|s\|_S = \|SP_G\|_S} \|G\Theta\left(\Phi[(\zeta s)^\intercal]_+\right)^\intercal\|_S \\
&= \sup_{s\, s.t. \|s\|_S = \|SP_G\|_S} \|G\Theta\left(\Phi(\zeta s)^\intercal\right)^\intercal\|_S
\end{aligned}$$

because to be norm-maximizing, applying the annihilator must not be norm-reducing (the norm-maximizing s is a signal lagged arbitrarily far into the past) which also means $(\zeta s)^\intercal = s^\intercal \zeta^\intercal + \tilde{\kappa}$ for an arbitrarily small $\tilde{\kappa}$. Using the supremum, the $\tilde{\kappa}$ can be ignored:

$$= \sup_{s\, s.t. \|s\|_S = \|SP_G\|_S} \|G\Theta\left(\Phi s^\intercal \zeta^\intercal\right)^\intercal\|_S$$

next, consider the norm-maximizing object $z = \Phi s^\intercal$ instead of s alone. The supremum becomes:

$$\begin{aligned}
= \sup_{z\, s.t.\ \|z\|_S = \|\Phi s^\intercal\|_S} \|G\Theta\left(z^\intercal \zeta^\intercal\right)^\intercal\|_S &= \sup_{z\, s.t.\ \|z\|_S = \|\Phi s^\intercal\|_S} \|G\Theta\zeta z\|_S \\
&\leq \sup_{z\, s.t.\ \|z\|_S = \|\Phi\|\|s^\intercal\|_S} \|G\Theta\zeta z\|_S
\end{aligned}$$

because $\|\Phi s^\intercal\|_S \leq \|\Phi\|\|P_G S^\intercal\|_S$ by definition of the operator norm and the constraint $\|s^\intercal\|_S = \|SP_G\|_S$. The operator norm definition also implies this quantity is

$$= \|G\Theta\zeta\|\|\Phi\|\|SP_G\|_S$$

Second, consider the idiosyncratic term, which does not simplify so nicely. In the image of $\mathcal{B}$, $S(I - P_G) = S_X(I - P_G)$:

$$\begin{aligned}
&\|G\Theta\left(\Phi[(\zeta S(I-P_G))^\intercal]_+\right)^\intercal\|_S \\
&\qquad = \|G\Theta\left(\Phi[(\zeta S_X(I-P_G))^\intercal]_+\right)^\intercal\|_S \\
&\qquad \leq \|G\Theta\|\|\left(\Phi[(\zeta S_X(I-P_G))^\intercal]_+\right)^\intercal\|_S \\
&\qquad = \|G\Theta\|\|\Phi[(\zeta S_X(I-P_G))^\intercal]_+\|_S \\
&\qquad \leq \|G\Theta\|\|\Phi\|\|[(\zeta S_X(I-P_G))^\intercal]_+\|_S \\
&\qquad = \|\Phi\|\vartheta_I
\end{aligned}$$

per the definition $\vartheta_I = \|G\Theta\| \| \left[\zeta S_X(I - P_G)\right]_+ \|_S$, and that the signal norm is block-transpose invariant.

Substitute the aggregate and idiosyncratic component bounds back into inequality (64)

$$\|G\Theta \left(\Phi[(\zeta S)^\intercal]_+\right)^\intercal \|_S \leq \|\Phi\| \|G\Theta\zeta\| \|SP_G\|_S + \|\Phi\|\vartheta_I$$

■

Lemma 10 *Let X and Y be orthogonal signal processes, and let $\lambda_X X + \lambda_Y Y$ be a linear combination with norm z:*

$$z = \|\lambda_X X + \lambda_Y Y\|_S$$

For scalars a and b, the linear combination that maximizes the quantity $a\|\lambda_X X\|_S + b\|\lambda_Y Y\|_S$ gives the value

$$\max_{\lambda_X, \lambda_y \ s.t. \ \|\lambda_X X + \lambda_Y Y\|_S = z} a\|\lambda_X X\|_S + b\|\lambda_Y Y\|_S = z\sqrt{a^2 + b^2}$$

Proof. X and Y are orthogonal, so the square of the norm of the linear combination is

$$z^2 = \|\lambda_X X + \lambda_Y Y\|_S^2 = \|\lambda_X X\|_S^2 + \|\lambda_Y Y\|_S^2$$

Define the new variables x and y such that $x = \frac{\lambda_X}{z}\|X\|_S$ and $y = \frac{\lambda_Y}{z}\|Y\|_S$. In these variables, the constrained maximization problem is

$$\max_{x,y} azx + bzy$$

$$s.t. \ \ 1 = x^2 + y^2$$

Rewritten as a maximization problem in one variable:

$$\max_x azx + bz\sqrt{1 - x^2}$$

has first order condition

$$a = b\frac{x}{\sqrt{1 - x^2}}$$

$$a^2(1 - x^2) = b^2 x^2$$

$$\frac{a^2}{b^2 + B^2} = x^2$$

and

$$y^2 = 1 - x^2 = \frac{b^2}{a^2 + b^2}$$

Taking roots, the solutions for x and y are

$$x = \frac{a}{\sqrt{a^2 + b^2}} \qquad\qquad y = \frac{b}{\sqrt{a^2 + b^2}}$$

Plug these values back into the objective function:

$$azx + bzy = z\frac{a^2}{\sqrt{a^2+b^2}} + z\frac{b^2}{\sqrt{a^2+b^2}} = z\sqrt{a^2+b^2}$$

■

A.9 Proofs for the Modified Confounding Dynamics Model with Idiosyncratic Noise

Proof of Proposition 4. Conjecture that $p_t = bw_t^x$ is an equilibrium price process for some b. Then agents' individual price $p_{i,t} = \beta E_{i,t}[x_{t+1}]$ is given by

$$p_{i,t} = \beta E[x_{t+1}|\{z_{i,t-j}, s_{i,t-j}\}_{j=0}^{\infty}] = \beta E[\alpha u_t|\{z_{i,t-j}, s_{i,t-j}\}_{j=0}^{\infty}] = \beta E[\alpha u_t|w_{i,t}^z, s_{i,t}] \quad (65)$$

where $w_{i,t}^z$ denotes the individual Wold innovation to signal $z_{i,t}$, given by

$$w_{i,t}^z = \frac{1}{1+\theta L} z_{i,t} = w_t^x + \tau_y^{-\frac{1}{2}} \epsilon_{i,t}^y$$

With this structure, equation (65) becomes

$$p_{i,t} = \beta E[w_t^x|w_{i,t}^z, s_{i,t}] = \beta \frac{\tau_y w_{i,t}^z + \tau_v b s_{i,t}}{\tau_y + \tau_v b^2 + 1}$$

per Lemma 12. Therefore the aggregate price follows

$$p_t = \beta \frac{\tau_y + \tau_v b}{\tau_y + \tau_v b^2 + 1} w_t^x$$

and the conjecture $p_t = bw_t^x$ requires that b solves the cubic

$$b(\tau_y + \tau_v b^2 + 1) = \beta(\tau_y + \tau_v b) \quad (66)$$

The discriminant D associated with the cubic equation is

$$D = -4\tau_v(\tau_y + 1 - \beta\tau_v)^3 - 27\tau_v^2\beta^2\tau_y^2$$

which is negative if and only if

$$27\beta^2\tau_v\tau_y^2 > 4(\beta\tau_v - \tau_y - 1)^3$$

implying a unique b solves equation (66).

■

Proof of Proposition 5. Per Proposition 2, in this model

$$\|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\| = \beta \quad (67)$$

and this operator is not affected by modifying the exogenous signal S_X to include idiosyncratic shocks.

The endogenous signal radius $R_N = \frac{\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|\|S_X P_G\|_S+\vartheta_I}{1-\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|}$ depends on the signal norm of the exogenous process S_X. In this model, the signal vector is

$$A_{i,t} = \begin{pmatrix} z_{i,t} \\ s_{i,t} \end{pmatrix} = \begin{pmatrix} (1+\alpha L)u_t + (1+\theta_L)\tau_y^{-\frac{1}{2}}\epsilon_{i,t}^y \\ p_t + \tau_v^{-\frac{1}{2}}\epsilon_{i,t}^v \end{pmatrix}$$

Only the price p_t is endogenous, so the exogenous process is

$$S_X = \begin{pmatrix} (1+\alpha L) & (1+\theta L)\tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix}$$

where the columns correspond to $(u_t, \epsilon_{i,t}^y, \epsilon_{i,t}^v)$ respectively, and $\tau_y^{-1/2}$ and $\tau_v^{-1/2}$ are the standard deviations of the idiosyncratic shocks. Per equation (19), the S_X operator has the block Toeplitz representation:

$$= \begin{pmatrix} \begin{pmatrix} 1 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \cdots \\ \begin{pmatrix} \alpha & \theta\tau_y^{-1/2} & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \cdots \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} \alpha & \theta\tau_y^{-1/2} & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 1 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

while the aggregate component zeroes any coefficient on idiosyncratic shocks: $S_X P_G = \begin{pmatrix} (1+\alpha L) & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$

Per Definition 2, the signal norm is the root sum of squared vector norms of the first three columns. For the aggregate component, only the first column is nonzero:

$$\|S_X P_G\|_S = \| \begin{pmatrix} 1 \\ \alpha \end{pmatrix} \| = \sqrt{1+\alpha^2} \tag{68}$$

The term $\vartheta_I = \|G\Theta\| \| \left[\Xi(B_{A1}L^{-1}+B_{A0})S_X(I-P_G)\right]_+ \|_S$ depends on the norm of a causal operator $G\Theta$. As explained in Section 2.3.2, this operator is simply $G\Theta = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, so $\|G\Theta\| = 1$. The non-causal operator is $\Xi(B_{A1}L^{-1}+B_{A0}) = L^{-1}\begin{pmatrix} \beta & 0 \end{pmatrix}$ so the forecasted idiosyncratic component is

$$\left[\Xi(B_{A1}L^{-1}+B_{A0})S_X(I-P_G)\right]_+ = \left[L^{-1}\begin{pmatrix} \beta & 0 \end{pmatrix}\begin{pmatrix} 0 & (1+\theta L)\tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix}\right]_+$$

$$= \left[L^{-1} \begin{pmatrix} 0 & \beta(1+\theta L)\tau_y^{-1/2} & 0 \end{pmatrix} \right]_+ = \begin{pmatrix} 0 & \beta\theta\tau_y^{-1/2} & 0 \end{pmatrix}$$

with norm $\| \left[\Xi(B_{A1}L^{-1} + B_{A0})S_X(I - P_G) \right]_+ \|_S = \beta\theta\tau_y^{-1/2}$. Putting these pieces together, φ_I is given by

$$\varphi_I = \beta\theta\tau_y^{-1/2} \tag{69}$$

Equations (67) (68) and (69) imply that the R_N bound on endogenous signals is

$$R_N = \frac{\beta\sqrt{1+\alpha^2} + \beta\theta\tau_y^{-1/2}}{1-\beta} \tag{70}$$

The SIC also depends on Σ_I, the variance of the contemporaneous idiosyncratic shocks. In this model $S_{X,0} = \begin{pmatrix} 1 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix}$, and the idiosyncratic component is

$$S_{X,0}(I - P_G) = \begin{pmatrix} 1 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \tau_y^{-1/2} & 0 \\ 0 & 0 & \tau_v^{-1/2} \end{pmatrix}$$

Therefore the variance of the contemporaneous idiosyncratic component is

$$\Sigma_I = \begin{pmatrix} \tau_y^{-1} & 0 \\ 0 & \tau_v^{-1} \end{pmatrix}$$

and the spectral radius of the inverse is

$$\mathrm{r}(\Sigma_I^{-1}) = \max\{\tau_y, \tau_u\} \tag{71}$$

Equations (67) (70) and (71) imply that the SIC holds if

$$\max\{\tau_y, \tau_u\} < \frac{1-\beta^2}{4\left(\frac{\beta\sqrt{1+\alpha^2}+\beta\theta\tau_y^{-1/2}}{1-\beta}\right)^2}$$

$$= \frac{(1-\beta^2)(1-\beta)^2}{4\beta^2\left(\sqrt{1+\alpha^2}+\theta\tau_y^{-1/2}\right)^2} = \frac{(1+\beta)(1-\beta)^3}{4\beta^2\left(\sqrt{1+\alpha^2}+\theta\tau_y^{-1/2}\right)^2}$$

■

A.10 Self-Map Lemma

Lemma 11 *$\mathcal{B}$ is an operator mapping $\mathcal{S}_{m_A,m_\varepsilon} \to \mathcal{S}_{m_A,m_\varepsilon}$*

Proof. The elements of $\mathcal{B}$ are in the following Banach spaces:

- $S_X \in \mathcal{S}_{m_A,m_\varepsilon}$
- $A \in \mathcal{S}_{m,m}$
- $X \in \mathcal{S}_{m,m}$
- $W \in \mathcal{S}_{m_A,m_\varepsilon}$
- $P_G \in \mathcal{S}_{n,n}$

The blocks agree so that $GXWP_G \in \mathcal{S}$, which is in the same space as S_X, so $S_X + GXWP_G \in \mathcal{S}$. ∎

A.11 Proofs of Propositions for the Example Models

A.11.1 Confounding Dynamics Proofs

This section proves several propositions about the confounding dynamics model introduced in Section 2.3.

Proof of Proposition 1. The forecast conditional on the confounding dynamics signal process is

$$E[x_{t+1}|\{p_{t-j}^{CD}, z_{i,t-j}\}_{j=0}^{\infty}] = E[x_{t+1}|\{w_{t-j}^F, w_{i,t-j}\}_{j=0}^{\infty}]$$

because the w_t^F process is invertible from the p_t^{CD} process, and the component of $z_{i,t}$ that is orthogonal to the w_t^F process is spanned by the idiosyncratic shock process $y_{i,t}$. Then, the expectation places no weight on the idiosyncratic shock because x_t is entirely aggregate:

$$= E[x_{t+1}|\{w_{t-j}^F\}_{j=0}^{\infty}] = [L^{-1}A^F(L)]_+ w_t^F$$

Thus with this information structure, the equilibrium price is

$$p_{i,t} = \beta E[x_{t+1}|\{w_{t-j}^F\}_{j=0}^{\infty}] = p_t^{CD}$$

∎

Proof of Proposition 2. Information Feedback Regularity is satisfied in this model: the operator $G\Theta\Xi(B_{A1}L^{-1} + B_{A0}) = \begin{pmatrix} 0 & 0 \\ L^{-1}\beta & 0 \end{pmatrix}$ has norm $\beta < 1$, because $L^{-1}\beta$ is the only non-zero entry, and $\|L^{-1}\beta\| = \beta\|L^{-1}\| = \beta$. ∎

Proof of Proposition 3. The shock vector $\varepsilon_{i,t} = \begin{pmatrix} u_t \\ y_{i,t} \end{pmatrix}$ is revealed by the time series $A_{i,t} = \begin{pmatrix} z_{i,t} \\ p_t \end{pmatrix}$ because $A_{i,t} = A(L)\varepsilon_{i,t}$ and $A(L)$ is invertible for the full information solution.

Consider a deviation $A^{\Delta}_{i,t} = A^{\Delta}(L)\varepsilon_{i,t}$ in a ball around $A_{i,t}$ with radius Δ such that $\|A^{\Delta} - A\| < \Delta$. The set of square invertible operators is open, so there exists a radius Δ such that all deviations $A^{\Delta}(L)$ are invertible.

Next, consider the signal operator of any such deviation $\mathcal{B}(S^{\Delta})$, and note that the signal operator S^{Δ} is equivalent to the Wold representation A^{Δ} because the signal is invertible. Equation (18) implies that the deviation in signal operators is given by

$$\begin{aligned}\mathcal{B}(A^{\Delta}) - \mathcal{B}(A) &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left[L^{-1} \begin{pmatrix} \beta & 0 \end{pmatrix} A^{\Delta}\right]_+ P_G - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left[L^{-1} \begin{pmatrix} \beta & 0 \end{pmatrix} A\right]_+ P_G \\ &= \left[L^{-1} \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \left(A^{\Delta} - A\right)\right]_+ P_G\end{aligned}$$

Take signal norms:

$$\begin{aligned}\|\mathcal{B}(A^{\Delta}) - \mathcal{B}(A)\|_S &= \| \left[L^{-1} \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \left(A^{\Delta} - A\right)\right]_+ P_G\|_S \\ &\leq \|L^{-1} \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \left(A^{\Delta} - A\right) \|_S\end{aligned}$$

because $[\cdot]_+$ and P_G are projections

$$\leq \|L^{-1} \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \|\| \left(A^{\Delta} - A\right) \|_S$$

by definition of the operator norm. Finally, $\|L^{-1} \begin{pmatrix} 0 & 0 \\ \beta & 0 \end{pmatrix} \| = \beta < 1$, so it must be that

$$\|\mathcal{B}(A^{\Delta}) - \mathcal{B}(A)\|_S < \| \left(A^{\Delta} - A\right) \|_S$$

and the full information solution $A_{i,t}$ must be signal-stable. ■

A.11.2 Beauty Contests

This section proves results about the beauty contests studied in Section 5.1.

Proof of Proposition 6. The operator $\begin{pmatrix} 0 & \varphi & \alpha L^{-1} \\ 0 & 0 & 0 \\ 0 & L\varphi & \alpha \end{pmatrix}$ has the block Toeplitz representation $\mathbf{C}$ per equation (19):

$$\mathbf{C} = \left(\begin{array}{cccc} \begin{pmatrix} 0 & \varphi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \alpha \end{pmatrix} & \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \cdots \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \varphi & 0 \end{pmatrix} & \begin{pmatrix} 0 & \varphi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \alpha \end{pmatrix} & \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \cdots \\ \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \varphi & 0 \end{pmatrix} & \begin{pmatrix} 0 & \varphi & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \alpha \end{pmatrix} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{array} \right)$$

and iterating the operator n times gives

$$\begin{pmatrix} 0 & \varphi & \alpha L^{-1} \\ 0 & 0 & 0 \\ 0 & L\varphi & \alpha \end{pmatrix}^n = \begin{pmatrix} 0 & \varphi\alpha^{n-1} & \alpha^n L^{-1} \\ 0 & 0 & 0 \\ 0 & L\varphi\alpha^{n-1} & \alpha^n \end{pmatrix} \implies \mathbf{C}^n = \alpha^{n-1}\mathbf{C}$$

What is the norm of this operator? Let $\mathbf{C}_i$ denote the ith column of $\mathbf{C}$. Then by definition:

$$\|\mathbf{C}\| = \max_{w_i,\ s.t.\ \sum_{i=1}^{\infty} w_i^2 = 1} \left(\sum_{i=1}^{\infty} \|w_i \mathbf{C}_i\|^2 \right)^{1/2}$$

which is bounded below by maximum row and column norms. These are the columns with either two φ terms or two α terms, or the rows with a single term of each:

$$\|\mathbf{C}\| \geq \max \left\{ \| \begin{pmatrix} \varphi \\ \varphi \end{pmatrix} \|, \| \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \|, \| \begin{pmatrix} \varphi & \alpha \end{pmatrix} \| \right\}$$

$$= \max \left\{ \sqrt{2}\varphi, \sqrt{2}\alpha, \sqrt{\varphi^2 + \alpha^2} \right\}$$

Thus IFR is satisfied only if $\max\left\{\sqrt{2}\varphi, \sqrt{2}\alpha, \varphi^2 + \alpha^2\right\} < 1$.

The spectral radius is even simpler:

$$\mathrm{r}\left(\mathbf{C}\right) = \lim_{n\to\infty} \|\mathbf{C}^n\|^{\frac{1}{n}} = \lim_{n\to\infty} \left(\|\mathbf{C}\|\alpha^{n-1}\right)^{\frac{1}{n}} = \alpha$$

Thus the Information Feedback Sub-Regularity condition (Appendix B) is satisfied only if $\alpha < 1$. ■

Lemma 12 *For two noisy signals s_1 and s_2 of a random variable $x \sim N(0,1)$ given by*

$$s_1 = \beta_1 x + u_1 \qquad\qquad s_2 = \beta_2 x + u_2$$

where $u_1 \sim N(0, \tau_1^{-1})$, $u_2 \sim N(0, \tau_2^{-1})$, and u_1, u_2, and x are all uncorrelated. The conditional expectation of x is given by

$$E[x|s_1, s_2] = \frac{\beta_1\tau_1 s_1 + \beta_2\tau_2 s_2}{\beta_1^2\tau_1 + \beta_2^2\tau_2 + 1}$$

Proof. The random variables are normal, so the expectation is linear:

$$E[x|s_1, s_2] = b_1 s_1 + b_2 s_2$$

for some b_1 and b_2. Per the law of total expectation:

$$Cov(x, s_1) = b_1 Var(s_1) + b_2 Cov(s_2, s_1) \qquad Cov(x, s_2) = b_1 Cov(s_1, s_2) + b_2 Var(s_2)$$

The variances and covariances imply

$$\beta_1 = b_1(\beta_1^2 + \tau_1^{-1}) + b_2\beta_1\beta_2 \qquad \beta_2 = b_1\beta_1\beta_2 + b_2(\beta_2^2 + \tau_2^{-1})$$

Solving the system gives

$$b_1 = \frac{\beta_1\tau_1}{\beta_1^2\tau_1 + \beta_2^2\tau_2 + 1} \qquad b_2 = \frac{\beta_2\tau_2}{\beta_1^2\tau_1 + \beta_2^2\tau_2 + 1}$$

■

A.12 Exponential Stability Proofs

This section proves results related to exponentially stable fixed points (Appendix B).

Proof of Proposition 7. Information Feedback Sub-Regularity is satisfied in this model: the relevant operator is represented as

$$G\Theta\Xi(B_{A1}L^{-1} + B_{A0}) = \begin{pmatrix} 0 & 0 \\ L^{-1}\beta & 0 \end{pmatrix}$$

so for any power $n > 1$, $\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n = 0$. Therefore the spectral radius is

$$\mathrm{r}\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right) = \lim_{n\to\infty} \| \left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)^n \|^{\frac{1}{n}} = 0$$

■

Proof of Theorem 9. Apply Lemma 5, raising both sides of equation (50) to $1/n$ and taking the limit implies that the spectral radii satisfy

$$\mathrm{r}\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right) = \mathrm{r}\left(\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_S}\mathbf{L}_{P_S}^{m_A}\right)$$

With this equality, Lemma 13 says the spectral radius of the Fréchet derivative $\mathrm{r}\left(D_{\mathcal{B}}(\hat{S})\right)$ at a fixed point $\hat{S}$ satisfies

$$\mathrm{r}\left(D_{\mathcal{B}}(\hat{S})\right) \geq \mathrm{r}\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)$$

Thus IFR is necessary for $\mathrm{r}\left(D_{\mathcal{B}}(\hat{S})\right) < 1$ to hold at a fixed point, which by Property 4 is true if and only if the fixed point is exponentially stable. ■

Lemma 13 *The spectral radius of the Fréchet derivative $r\left(D_{\mathcal{B}}(\hat{S})\right)$ at a fixed point $\hat{S}$ satisfies*

$$r\left(D_{\mathcal{B}}(\hat{S})\right) \geq r\left(\mathbf{L}_{P_G}^{m_A}\mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon}\mathbf{Q}_{P_{\hat{S}}}\mathbf{L}_{P_{\hat{S}}}\right)$$

Proof.

Consider a signal with Wold representation $S = AW$. Let $\mathcal{S}_{P_S}$ denote the subset of the signal space that is spanned by current and past S (and thus W). Let $\mathcal{S}_{M_S}$ denote the residual space orthogonal to current and past S. The sets are orthogonal complements so that

$$\mathcal{S}_{m_A,m_\varepsilon} = \mathcal{S}_{P_S} \oplus \mathcal{S}_{M_S}$$

Let $\mathcal{B}P_S(\tilde{S})$ denote the operator projecting a signal $\tilde{S}$ onto current and past S (i.e. onto the space $\mathcal{S}_{P_S}$) and then applying $\mathcal{B}$.

First I characterize the image of $\mathcal{B}P_S$. The Wold representation $S = AW$ features invertible operator A. The set of square invertible operators is open, so there exists a radius δ such that all deviations $\tilde{A}^\Delta$ with $\|\tilde{A}^\Delta - A\|_S < \delta$ are invertible. Now consider a deviation S^Δ such that $\|S^\Delta - S\|_S < \delta$ and $S^\Delta \in \mathcal{S}_{P_S}$ so that the deviation is spanned by current and past W, i.e. $S^\Delta = A^\Delta W$ for some A^Δ. $\|A^\Delta - A\|_S = \|S^\Delta - S\|_S < \delta$, so A^Δ is invertible and $A^\Delta W$ is the Wold representation. Therefore $S^\Delta \in \mathcal{S}_{P_S}$. This implies that $\mathcal{B}P_S(\tilde{S})$ maps $\mathcal{S}_{P_S} \to \mathcal{S}_{P_S}$ in a neighborhood around S.

Next, consider a fixed point $\hat{S} = \mathcal{B}(\hat{S})$, and a deviation $\hat{S} + D$. $\mathcal{D}_{\mathcal{B}}(\hat{S})$ is the Fréchet derivative, so

$$\mathcal{B}(\hat{S}+D) - \mathcal{B}(\hat{S}) = \hat{S} + \mathcal{D}_{\mathcal{B}}(\hat{S})D + o(\|D\|_S) - \hat{S} = \mathcal{D}_{\mathcal{B}}(\hat{S})D + o(\|D\|_S)$$

which implies

$$\mathcal{B}^n(\hat{S}+D) - \mathcal{B}^n(\hat{S}) = \mathcal{B}^{n-1}(\hat{S} + \mathcal{D}_{\mathcal{B}}(\hat{S})D + o(\|D\|_S)) - \mathcal{B}^{n-1}(\hat{S})$$

and repeated iteration gives

$$\mathcal{B}^n(\hat{S}+D) - \mathcal{B}^n(\hat{S}) = \mathcal{D}_{\mathcal{B}}(\hat{S})^n D + o(\|D\|_S)$$

Because $\mathcal{B}$ maps $\mathcal{S}_{P_{\hat{S}}} \to \mathcal{S}_{P_{\hat{S}}}$ on a neighborhood around $\hat{S}$, the Fréchet derivative $\mathcal{D}_{\mathcal{B}}(\hat{S})$ also maps $\mathcal{S}_{P_{\hat{S}}} \to \mathcal{S}_{P_{\hat{S}}}$. This implies that

$$\mathcal{D}_{\mathcal{B}}(\hat{S})P_{\hat{S}} = P_{\hat{S}}\mathcal{D}_{\mathcal{B}}(\hat{S})P_{\hat{S}} \tag{72}$$

where again $P_{\hat{S}}$ denotes projecting onto current and past $\hat{S}$ before or after applying the linear operator $\mathcal{D}_{\mathcal{B}}(\hat{S})$. $M_{\hat{S}} = I - P_{\hat{S}}$ denotes projecting onto the orthogonal subspace.

Next consider the operator norm of nth power of the Fréchet derivative:

$$\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n\| = \|\mathcal{D}_{\mathcal{B}}(\hat{S})^n(P_{\hat{S}} + M_{\hat{S}})\|$$

From the operator norm definition, consider each term:

$$\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n P_{\hat{S}}\| = \sup_{y_{\hat{S}}} \frac{\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n P_{\hat{S}} y_{\hat{S}}\|_S}{\|y_{\hat{S}}\|_S} \qquad \|\mathcal{D}_{\mathcal{B}}(\hat{S})^n M_{\hat{S}}\| = \sup_{y_{\perp\hat{S}}} \frac{\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n M_{\hat{S}} y_{\perp\hat{S}}\|_S}{\|y_{\perp\hat{S}}\|_S}$$

The signals $y_{\hat{S}} \in \mathcal{S}_{P_{\hat{S}}}$ and $y_{\perp\hat{S}} \in \mathcal{S}_{M_{\hat{S}}}$ maximizing these norms are orthogonal; therefore, the unit signal $y = y_{\hat{S}} + y_{\perp\hat{S}}$ maximizing $\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n y\|_S$ is a linear combination satisfying $\|y_{\hat{S}}\|_S^2 + \|y_{\perp\hat{S}}\|_S^2 = 1$. By lemma 10, maximizing $\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n P_{\hat{S}} y_{\hat{S}} + \mathcal{D}_{\mathcal{B}}(\hat{S})^n M_{\hat{S}} y_{\perp\hat{S}}\|_S$ subject to this constraint implies the operator norm satisfies

$$\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n\| = \sqrt{\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n P_{\hat{S}}\|^2 + \|\mathcal{D}_{\mathcal{B}}(\hat{S})^n M_{\hat{S}}\|^2} \tag{73}$$

Combining equation (73) with equation (72) implies

$$\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n\| \geq \| \left(\mathcal{D}_{\mathcal{B}}(\hat{S}) P_{\hat{S}}\right)^n \|$$

$$\|\mathcal{D}_{\mathcal{B}}(\hat{S})^n\|^{\frac{1}{n}} \geq \| \left(\mathcal{D}_{\mathcal{B}}(\hat{S}) P_{\hat{S}}\right)^n \|^{\frac{1}{n}}$$

and in the limit

$$\mathrm{r}\left(D_{\mathcal{B}}(\hat{S})\right) \geq \mathrm{r}\left(\mathcal{D}_{\mathcal{B}}(\hat{S}) P_{\hat{S}}\right)$$

Finally, Theorem 8 gives the matrix representation for the Fréchet derivative acting on $\mathcal{S}_{P_{\hat{S}}}$ as $\mathbf{L}_{P_G}^{m_A} \mathbf{R}_{(G\Theta)^\intercal}^{m_\epsilon} \mathbf{Q}_{P_{\hat{S}}} \mathbf{L}_{P_{\hat{S}}}$. ■

B Exponential Stability

This section introduces an alternative type of stability: exponential stability. Many properties of signal-stable equilibria also apply to exponentially stable equilibria. The main drawback of studying this type of stability is that it has no analog to Theorem 6: exponentially stable fixed points may not be globally unique.

B.1 Definitions

The signal-stability property defined in Section 3.4 represents "contractive stability". Everywhere in a neighborhood around a signal-stable fixed point, $\mathcal{B}$ is a contraction.

Contractive stability is a strong property. A weaker form of stability is "exponential stability", where there is some k such that $\mathcal{B}^k$ is a contraction on a neighborhood around the fixed point. Under this definition, if there is a small perturbation to a fixed point, the operator $\mathcal{B}$ can cause large deviations in signals, but after enough repeated applications of $\mathcal{B}$, the signal must converge. Put another way, exponentially stable fixed points are robust to a single perturbation, but not necessarily repeated perturbations. This form of stability is formally defined as follows:

Definition 6 *An equilibrium fixed point signal satisfying $S = \mathcal{B}(S)$ is called* exponentially stable *if there exists some neighborhood of S, some $\alpha \in (0,1)$ and $\kappa \geq 1$ such that for any S^Δ in the neighborhood, $\|\mathcal{B}^k(S^\Delta) - \mathcal{B}^k(S)\|_S < \kappa\alpha^k \|S^\Delta - S\|_S$ for all $k \geq 0$.*

Because exponential stability is a weaker property than signal-stability, IFR is no longer a necessary condition. However, exponential stability does have an analogous condition: Information Feedback Sub-Regularity (IFSR).

Condition 3 *A model satisfies Information Feedback Sub-Regularity if*

$$r\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right) < 1$$

$\mathrm{r}(\cdot)$ here denotes the spectral radius, which measures by how much repeated application of an operator can increase the variance of any signal process. Specifically, for a linear operator A, the spectral radius is given by

$$\mathrm{r}(A) \equiv \lim_{k\to\infty} \|A^k\|^{\frac{1}{k}}$$

where $\|\cdot\|$ is the operator norm. The spectral radius is equivalent to the largest eigenvalue of a finite dimensional matrix. (Although in infinite dimensions, there may not exist a largest eigenvalue.)

Accordingly, Condition 3 says that repeated application of this operator must decrease the variance of a signal. The spectral radius of this operator represents how much $\mathcal{B}^k(S)$ (with k large) can be changed by perturbing the signal process S in a way that is spanned by the forecast error process W.

Proposition 3 demonstrated in the asset pricing example (Section 2.3) that Information Feedback Regularity is straightforward to check from a models assumptions. Proposition 7 shows that sub-regularity is similarly straightforward:

Proposition 7 *Information Feedback Sub-Regularity is always satisfied in the confounding dynamics model.*

Proof: Appendix A.12

This result reveals that exponential stability is weaker than signal stability. In the confounding dynamics model, $\beta < 1$ is necessary for signal-stable fixed points to exist. This is because there is a perturbation to the fixed point that increases the implied signal deviation by a factor of β. However, if this is a one-off perturbation, the deviation is short-lived. Proposition 7 implies that repeated application of $\mathcal{B}$ will return to the fixed point for any value of β.

B.2 Properties of Exponentially Stable Fixed Points

This section describes some properties of exponentially stable fixed points. Each is an analog a signal-stability property derived in Section 3.4.

The norm of the Fréchet derivative $D_{\mathcal{B}}(\hat{S})$ determined if a fixed point $\hat{S}$ was signal-stable. Similarly, the spectral radius of $D_{\mathcal{B}}(\hat{S})$ determines if the fixed point is exponentially stable:

Property 4 *$\hat{S}$ is an exponentially stable fixed point if and only if $\mathrm{r}\Big(D_{\mathcal{B}}(\hat{S})\Big) < 1$.*

Proof: Appendix A.12 This property follows from (Garab, Pituk, and Pötzsche, 2020, Thm. 1) by redefining $\hat{S}$ to be the origin. Theorem 8 gives the exact expression for $D_{\mathcal{B}}(\hat{S})$.

Theorem 9 is the analog to Theorem 3; Information Feedback Sub-Regularity is necessary for an exponentially stable fixed point to exist.

Theorem 9 *If all fixed points of a model contain aggregate signals such that for any fixed point signal vector $\hat{S}$ there is an entry $\hat{S}_i$ satisfying $\hat{S}_i P_G = \hat{S}_i$ then Information Feedback Sub-Regularity is a necessary condition for signal-stable fixed points to exist.*

Proof: Appendix A.12

As with signal-stability (Theorem 4), exponentially stable fixed points are locally unique, because there exists some k such that $\mathcal{B}^k$ is a contraction on a neighborhood around the fixed point. However, it is impossible to say in general if exponentially stable fixed points are globally unique. There is no exponential analog to Theorem 6 because there is no analog to Lemma 7: the set of signals for which $\|D_{\mathcal{B}}(S)\| < 1$ is path connected, but this is not true of the set of signals for which $\mathrm{r}\left(D_{\mathcal{B}}(S)\right) < 1$. Indeed, the beauty contest model in Section 5.1 is an example of this; in the multiplicity region (Figure 2), there are two exponentially stable fixed points separated by an unstable fixed point.

Fortunately, even if economists are *ex ante* interested in exponential stability, studying signal-stability is still useful for two reasons. First, any signal-stable equilibrium is necessarily exponentially stable, because the spectral radius of the Fréchet derivative is bounded by its norm:

$$\mathrm{r}\left(D_{\mathcal{B}}(S)\right) \leq \|D_{\mathcal{B}}(S)\|$$

Second, signal-stability was necessary to prove the global uniqueness result in Theorem 7, which says that if the Sufficient Idiosyncrasy Condition holds, then there exists a globally unique fixed point, and it will be signal-stable. Therefore, if SIC holds, then there exists a globally unique exponentially stable fixed point as well.

C Computation

This appendix formally introduces the Signal Operator Iteration algorithm and describes a method for computing it.

C.1 Theoretical Algorithm

The algorithm applying equation (22) is straightforward to describe informally. Begin by guessing a signal process $S^n(L)$. Then, find the policy function $X^n(L)$ implied by the signal process by inverting the signal to find the forecast errors $W^n(L)$ and applying the solution method from Section 2. Next, use the assumed relationship between endogenous variables and endogenous information that is encoded in $G(L)$ to calculate the implied signal process $S^{n+1}(L)$. Repeat until the signal process converges.

In practice, this algorithm can quickly become uncomputable: the signal is high dimensional, and the dimension may increase with every iteration of the algorithm. This is an unavoidable challenge, because the true equilibrium may be infinite-dimensional. Therefore an additional step is necessary to ensure the algorithm remains computable. A standard approach is known as the "finite section method" (Böttcher and Silbermann, 2012), which truncates a signal process after some fixed number of lags. I refer to this truncation length as the "order" of the algorithm, and the operator P_τ represents truncation after lag τ.

Appendix C.3 details how to compute this algorithm in practice. Formally, the algorithm is:

Algorithm 1 (Signal Operator Iteration) *Conjecture a square-summable causal lag operator polynomial $S^0(L)$. Then proceed with iteration $n = 0$ as follows:*

1. *Find the autocovariance function $\Gamma^n(L)$ implied by $S^n(L)$ using equation* (12)*.*

2. *Decompose $\Gamma^n(L)$ to find the forecast error process $W^n(L)$ and moving average representation $A^n(L)$*

3. *Calculate the policy function $X^n(L)$ from $A^n(L)$ by Theorem 1.*

4. *Construct the endogenous signal $S^n_N(L)$ by equation* (15)*:*

$$S^n_N(L) = [G(L)X^n(L)W^n(L)P_G]_+$$

5. *Calculate the next signal polynomial $S^{n+1}(L)$ by combining signals using equation* (10) *and truncating to order τ:*

$$S^{n+1}(L) = (S_X(L) + S^n_N(L))\,P_\tau \tag{74}$$

6. *If $\|S^{n+1} - S^n\|$ is sufficiently close to zero, conclude that $S(L) = S^{n+1}(L)$. Otherwise return to Step 1 with guess S^{n+1}.*

C.2 Properties of the Approximation

Approximating operators in this way allows for arbitrarily precise approximation of the solution $\hat{S}$. The signal Toeplitz operators S map $\ell^2 \to \ell^2$, implying the corresponding lag operator polynomials have square summable coefficients, and the infinite matrix features exponential decay off the main diagonal. When the operator S is approximated by an operator S^τ which has truncation length τ, Strohmer (2002) proves that the error to linear operations and inversion can be made arbitrarily small by choosing a large enough value of τ.

It is practical to select a large value of τ, given that the solution algorithm is not computationally intensive. Because the solution $\hat{S}$ must be square summable, a strategy for checking whether τ is large enough is to select a small bound $\bar{b} > 0$ below which terms are considered sufficiently close to zero, and then check that all terms s in the τth block $\hat{S}_\tau$ of the computed solution are within the bounds, so that $|s| < \bar{b}$. If not, increase the truncation length τ.

This approximation method is well-suited for this problem specifically because the algorithm uses causal operators. Usually, approximating operators on infinite Toeplitz matrices also requires embedding into a circulant matrix, which introduces perturbation error in addition to the truncation error. This is because even though S is approximated by S^τ, S^τ is still an infinite matrix. However, causal operators have upper block triangular Toeplitz matrices, so it's possible to calculate the truncated product of two truncated Toeplitz matrices without any additional operators. Theorem 10 formalizes this property.

Suppose A, B, C are operators mapping $\ell^2 \to \ell^2$ with conformable blocks: the blocks of A are $k \times n$, the blocks of B are $k \times m$, and the blocks of C are $m \times n$. Let $T^\tau(A), T^\tau(B), T^\tau(C)$ denote the $\tau k \times \tau n$, $\tau k \times \tau m$, and $\tau m \times \tau n$ block Toeplitz matrices with the same main diagonal blocks as the infinite operators.

Theorem 10 *If operators A, B, C mapping $\ell^2 \to \ell^2$ are causal and satisfy*

$$A = BC$$

Then the finite approximations $T^\tau(A), T^\tau(B), T^\tau(C)$ satisfy

$$T^\tau(A) = T^\tau(B)T^\tau(C)$$

Proof. Partition the operators A, B, C into blocks of arbitrary but equal size. The equation $A = BC$ becomes

$$\begin{pmatrix} A_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ A_1 & A_0 & \mathbf{0} & \mathbf{0} & \cdots \\ A_2 & A_1 & A_0 & \mathbf{0} & \cdots \\ A_3 & A_2 & A_1 & A_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} B_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ B_1 & B_0 & \mathbf{0} & \mathbf{0} & \cdots \\ B_2 & B_1 & B_0 & \mathbf{0} & \cdots \\ B_3 & B_2 & B_1 & B_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} C_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ C_1 & C_0 & \mathbf{0} & \mathbf{0} & \cdots \\ C_2 & C_1 & C_0 & \mathbf{0} & \cdots \\ C_3 & C_2 & C_1 & C_0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

These matrices are block lower triangular, so the blocks A_0, B_0, C_0 satisfy

$$A_0 = B_0 C_0$$

If the operators A, B, C are partitioned into $\tau k \times \tau n$, $\tau k \times \tau m$, and $\tau m \times \tau n$ blocks respectively, then $T^\tau(A), T^\tau(B), T^\tau(C)$ appear on the main block diagonals. Therefore, they must satisfy

$$T^\tau(A) = T^\tau(B) T^\tau(C)$$

■

C.3 Computing the Algorithm

To compute the Signal Operator Iteration algorithm with finite Toeplitz approximations, I use the following steps. Begin then by conjecturing a causal square-summable signal process S^0 which is approximated by the finite block Toeplitz matrix $T^\tau(S^0)$. Then proceed by:

1. Find the autocovariance's finite block Toeplitz approximation implied by signal process S^n using using equation (12). For $j \in [-\tau, \tau]$, the blocks in the $T^\tau(\Gamma^n)$ block Toeplitz matrix are given by

$$\Gamma_j = \sum_{k=0}^{\tau} S_k^n {S_{k+j}^n}'$$

2. Use $T^\tau(\Gamma^n)$ to find the Wold representation: calculate $T^\tau(A^n)$ using one of the methods in Appendix C.4.

3. Given the Wold representation $T^\tau(A^n)$, generate the matrix $T^\tau(\tilde{A}^n)$ by equation (31). If $T^\tau(L^{-1})$ is the finite approximation to the inverse lag operator (i.e. a block matrix with identity matrices along the first block above the main diagonal and zeros elsewhere) and if $T(B_{A1})$ is the block matrix with B_{A1} along the main diagonal (and similarly for B_{A0}) then $T^\tau(\tilde{A}^n)$ is given by

$$T^\tau(\tilde{A}^n) = \left[\left(T^\tau(B_{A1}) T^\tau(L^{-1}) + T^\tau(B_{A0})\right) T^\tau(A^n)\right]_{LT}$$

   where the operator $[\cdot]_{LT}$ is the finite matrix equivalent of the annihilation operator $[\cdot]_+$ setting all blocks above the main diagonal to zero.

4. Calculate the block Toeplitz approximation of the policy function $T^\tau(X^n)$ by applying Theorem 1:

$$T^\tau(X^n) = T^\tau(\Theta) \left[T^\tau(\Xi) T^\tau(\tilde{A}^n)\right]_{LT}$$

5. Calculate the implied approximation of the signal $T^\tau(S^{n+1})$ using equation (74):

$$T^\tau(S^{n+1}) = T^\tau(S_X) + [T^\tau(G)T^\tau(X^n)T^\tau(A^n)^{-1}(L)T^\tau(S^n)T^\tau(P_G)]_{LT}$$

6. If the Euclidean matrix norm of $\|T^\tau(S^{n+1}) - T^\tau(S^n)\|_2$ is sufficiently close to zero, conclude that the equilibrium signal process is $S(L) = S^{n+1}(L)$. Otherwise return to Step 1 with guess S^{n+1}.

The finite matrix approximation introduces some error into this algorithm, although this error can be reduced by choosing an arbitrarily large approximation length τ. This is practical even for large values of τ because only the matrix inversion in Step (2.) is computationally intensive; the other steps are linear matrix operations. Concatenation error occurs in Steps (1.) and (2.), but goes to zero as τ becomes large. Theorem 10 ensures that the remaining steps introduce no additional error.

C.4 Computing the Wold Representation

How can the Wold representation $A(L)W(L) = S(L)$ be calculated? The innovation polynomial $A(L)$ and the signal polynomial $S(L)$ both produce the same series of $A_{i,t}$, so they must have the same autocovariance function. This sections describes two methods to calculate the polynomial $A(L)$. Then the white noise polynomial $W(L)$ can be found by $W(L) = A^{-1}(L)S(L)$.

C.4.1 Cholesky Method

Autocovariances are related to the Wold coefficients by the operator equation:

$$\begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ ... & \Gamma_0 & \Gamma_1 & \Gamma_2 & ... \\ ... & \Gamma_1 & \Gamma_0 & \Gamma_1 & ... \\ ... & \Gamma_2 & \Gamma_1 & \Gamma_0 & ... \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = L(AC_W)L(AC_W)' \tag{75}$$

where the bi-infinite Laurent operator $L(AC_W)$ is given by

$$L(AC_W) \equiv \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ ... & A_0C_W & 0 & 0 & ... \\ ... & A_1C_W & A_0C_W & 0 & ... \\ ... & A_2C_W & A_1C_W & A_0C_W & ... \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and where C_W denotes the Cholesky decomposition $\Sigma_W = C_W C_W'$.

These are infinite-dimensional operators, but can be approximated with finite block-Toeplitz matrices. In particular, Caines and Gerencser (1991) prove that $T^\tau(A)$ calculated from $T^\tau(\Gamma)$ by Cholesky decomposition converges to the true Wold representation as τ becomes large.

C.4.2 Yule-Walker Method

Alternatively, the Wold representation can be calculated using the autoregressive representation instead of the moving average representation, i.e. calculating A^{-1} instead of A from the autocovariances.

The innovation polynomial $A(L)$ is the Wold decomposition of the signal polynomial $S(L)$, so its inverse $A(L)^{-1}$ solves the Yule-Walker Equations:

$$\begin{pmatrix} \Gamma_0 & \Gamma_1 & \Gamma_2 & \dots \\ \Gamma_1 & \Gamma_0 & \Gamma_1 & \dots \\ \Gamma_2 & \Gamma_1 & \Gamma_0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} -(A^{-1})_1' \\ -(A^{-1})_2' \\ -(A^{-1})_3' \\ \vdots \end{pmatrix} = \begin{pmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \vdots \end{pmatrix} \tag{76}$$

where the polynomial $A(L)$ is normalized so that $A_0 = I$.

To calculate A^{-1}, use the finite $T^\tau(\Gamma)$ and calculate $T^\tau(A)^{-1}$ that solves the first τ Yule-Walker equations (76), and invert to find the MA representation $T^\tau(A)$.

Evaluating the Information Feedback Regularity Condition 1 requires calculating the norm $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|$. Sometimes this norm can be calculated analytically (e.g. the Singleton model in Section 5.2) but in most cases it must be calculated numerically. This section demonstrates that the approximation by finite section method can be made arbitrarily precise by choosing a large enough truncation order τ.

The norm of the τ-order approximation $\|T^\tau\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)\|$ converges to the true norm. This is not true for all properties of an operator (e.g. its trace), but the norms of truncated Toeplitz operators are known to converge (Böttcher and Silbermann, 2012):

$$\lim_{\tau\to\infty} \|T^\tau\left(G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\right)\| = \|G\Theta\Xi(B_{A1}L^{-1} + B_{A0})\|$$

C.5 Computing the Regularity Condition

In the associated programming package, the subroutine `ifrnorm` determines whether a model satisfies IFR.

Evaluating the Information Feedback Regularity Condition 1 requires calculating the norm $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|$. Sometimes this norm can be calculated analytically (e.g. the Singleton model in Section 5.2) but in most cases it must be calculated numerically. Fortunately, calculating this norm by the finite section method can be made arbitrarily precise by choosing a large enough truncation order τ.

The method is simple. First, write the operator $G\Theta\Xi(B_{A1}L^{-1}+B_{A0})$ as a product of large block Toeplitz matrices $T^\tau(G)T^\tau(\Theta)T^\tau(\Xi)T^\tau(B_{A1}L^{-1}+B_{A0})$. The proof of Proposition 6 gives a concrete example of step. Second, tell a computer to calculate the matrix norm $\|T^\tau(G)T^\tau(\Theta)T^\tau(\Xi)T^\tau(B_{A1}L^{-1}+B_{A0})\|$.

The norm of the τ-order approximation $\|T^\tau(G)T^\tau(\Theta)T^\tau(\Xi)T^\tau(B_{A1}L^{-1}+B_{A0})\|$ converges to the true norm. This is not true for all properties of an operator (e.g. its trace), but the norms of truncated Toeplitz operators are known to converge (Böttcher and Silbermann, 2012) so long as $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|$ is finite. The finiteness condition is relevant: there is a the unit root in the Section 5.3 Ξ operator, so numerical norms will grow with τ because $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\| = \infty$ in that example.

C.6 Computing the Fréchet Derivative Norm

In the associated programming package, the subroutine `soifrechet` calculates the norm of the derivative of a signal operator at a point.

By Corollary 2, the norm of the Fréchet Derivative is

$$\|D_{\mathcal{B}}(S)\| = \sqrt{\|\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{P_S}\mathbf{L}_{P_S}\|^2 + \|\mathbf{L}^{m_A}_{P_G}\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}\mathbf{Q}_{M_S}\mathbf{L}_{M_S}\|^2}$$

Both terms are linear operators. The first term is equivalent to $\|G\Theta\Xi(B_{A1}L^{-1}+B_{A0})\|$ (see the proof of Theorem 3) so it is calculated with `ifrnorm` as described in Appendix C.4. Similarly, `soifrechet` computes the norm of the second term by the finite section method. It constructs each linear operator $\mathbf{L}^{m_A}_{P_G}$, $\mathbf{R}^{m_\epsilon}_{(G\Theta)^\intercal}$ and so forth as a large block Toeplitz matrix, multiplies them, and then computes the matrix norm.

D Additional Stability Results in the Confounding Dynamics Model

Consider the following version of the confounding dynamics model introduced in Section 2.3. The fundamental value of the asset is given by

$$x_t = u_t + \alpha u_{t-1}$$

where $\alpha > 1$ and $Var(u_t) = 1$. The full information equilibrium of this model is

$$p_{i,t} = \beta E_{i,t}[\alpha u_t]$$

I use this example to demonstrate two properties: the full information equilibrium is signal-unstable if $\beta > 1$, and the confounding dynamics equilibrium is signal-unstable even if $\beta < 1$.

The signal vector in this model is $A_{i,t} = \begin{pmatrix} z_{i,t} \\ p_t \end{pmatrix}$. To demonstrate instability, it needs to be shown that

$$\|\mathcal{B}(A^{\Delta}) - \mathcal{B}(A)\|_S > \|A^{\Delta} - A\|_S$$

if A is the signal operator for any equilibrium, and A^{Δ} denotes an arbitrarily small perturbation of the signal operator.

D.1 Instability with Full Information

Theorem 3 implies that if $\beta > 1$, the full information equilibrium must be signal-unstable. This is easily demonstrated in the example, by perturbing the object of the pricing equation, $E_{i,t}[x^{\Delta}_{t+1}]$. The model's operator representation forecasts future noisy signals, so this change is encoded by perturbing the noisy signal. Therefore, consider the perturbed signal process $A^{\Delta}_{i,t}$:

$$A^{\Delta}_{i,t} = \begin{pmatrix} z_{i,t} + \Delta u_{t-1} \\ p^{FI}_t \end{pmatrix}$$

where $p^{FI}_t = \beta\alpha u_t$ denotes the full information equilibrium price process, and Δ is an arbitrary scalar.

The implied price for agents observing this perturbed signal process is

$$p^{\Delta}_t = \beta E[(\alpha + \Delta)u_t | \{p^{FI}_{t-j}\}^{\infty}_{j=0}]$$

$$= \beta E[(\alpha + \Delta)u_t | \{u_{t-j}\}^{\infty}_{j=0}] = \beta(\alpha + \Delta)u_t$$

The signal norms are the sums of standard deviations of the difference in signals. The initial perturbation is:

$$\|A^{\Delta} - A\|_S = \sqrt{Var(z_{i,t} + \Delta u_{t-1} - z_{i,t})} + \sqrt{Var(p^{FI}_t - p^{FI}_t)} = \Delta$$

and after the $\mathcal{B}$ operators are applied:

$$\|\mathcal{B}(A^{\Delta}) - \mathcal{B}(A)\|_S = \sqrt{Var(z_{i,t} - z_{i,t})} + \sqrt{Var(p^{\Delta}_t - p^{FI}_t)}$$

$$= \sqrt{Var(\beta(\alpha + \Delta)u_t - \beta\alpha u_t)} = \beta\Delta$$

Therefore, $\|\mathcal{B}(A^{\Delta}) - \mathcal{B}(A)\|_S > \|A^{\Delta} - A\|_S$ if $\beta > 1$, so the full information equilibrium is signal-unstable.

D.2 Instability with Confounding Dynamics

Proposition 3 and Theorem 6 imply that any confounding dynamics equilibrium must be signal-unstable, even when $\beta < 1$. In contrast to the last section, I will demonstrate instability by perturbing the observed price. Consider the perturbed signal process $A_{i,t}^{\Delta}$:

$$A_{i,t}^{\Delta} = \begin{pmatrix} z_{i,t} \\ p_t^{CD} + \frac{\Delta}{1+\theta L} u_t \end{pmatrix}$$

where $p_t^{CD} = \beta\theta w_t$ denotes the confounding dynamics equilibrium price process, Δ is an arbitrary scalar, and $\theta = \alpha^{-1}$. w_t is the forecast error process in the Wold representation of x_t:

$$x_t = w_t + \theta w_{t-1}$$

$$\implies w_t = \frac{1+\alpha L}{1+\theta L} u_t$$

In the math that follows, it is simpler to keep track of Blaschke factors instead of forecast errors. For example, define the Blaschke factor B^{CD} by

$$B^{CD} \equiv \frac{\theta + L}{1+\theta L}$$

which implies $B^{CD} u_t = \theta w_t$. This is helpful because Blaschke factors preserve variances, i.e. for any Blaschke factor B, $Var(Bu_t) = 1$.

In order to find the new price p_t^{Δ} implied by the perturbation, first I derive the Wold representation of the perturbed endogenous signal $p_t^{CD} + \frac{\Delta}{1+\theta L} u_t$:

$$p_t^{CD} + \frac{\Delta}{1+\theta L} u_t = \beta\theta w_t + \frac{\Delta}{1+\theta L} u_t$$

$$= \beta \frac{\theta + L}{1+\theta L} u_t + \frac{\Delta}{1+\theta L} u_t = \beta \frac{\theta + \frac{\Delta}{\beta} + L}{1+\theta L} u_t = \beta \frac{\xi + L}{1+\theta L} u_t$$

where $\xi \equiv \theta + \frac{\Delta}{\beta}$, which satisfies $\xi \in (0,1)$ for sufficiently small Δ

$$= \beta \frac{1+\xi L}{1+\theta L} \frac{\xi + L}{1+\xi L} u_t = \beta \frac{1+\xi L}{1+\theta L} B^{\Delta} u_t$$

for the Blaschke factor $B^{\Delta} \equiv \frac{\xi + L}{1+\xi L}$. $\beta \frac{1+\xi L}{1+\theta L}$ is invertible, so $B^{\Delta} u_t$ is proportional to the Wold representation's forecast error process.

Let $z_{i,t}^*$ denote the components of $z_{i,t}$ orthogonal to current and past $B^{\Delta} u_t$. Then the implied price is:

$$p_t^{\Delta} = \beta E_{i,t}[\alpha u_t] = \beta E[\alpha u_t | \{B^{\Delta} u_{t-j}, z_{i,t-j}^*\}_{j=0}^{\infty}]$$

$$= \beta E[\alpha u_t | \{B^{\Delta} u_{t-j}\}_{j=0}^{\infty}] + \beta E[\alpha u_t | \{z_{i,t-j}^*\}_{j=0}^{\infty}] = \beta E[\alpha u_t | \{B^{\Delta} u_{t-j}\}_{j=0}^{\infty}] + o(\Delta)$$

where the first step is implied by orthogonality, and the second step is implied by $\lim_{\Delta\to 0} cov(u_t, z^*_{i,t-j}) = 0$; as the perturbation goes to zero, $B^\Delta u_t \to B^{CD} u_t$ which spans the x_t component of z_t, but never the noisy $y_{i,t}$ component.

$$= \beta\alpha cov(u_t, B^\Delta u_t) B^\Delta u_t + o(\Delta) = \beta\alpha\xi B^\Delta u_t + o(\Delta)$$

Next, consider the norm of the perturbation from the confounding dynamics equilibrium. The initial perturbation is:

$$\|A^\Delta - A^{CD}\|_S = \sqrt{Var(z_{i,t} - z_{i,t})} + \sqrt{Var(p^{CD}_t + \frac{\Delta}{1+\theta L} u_t - p^{CD}_t)} = \frac{\Delta}{\sqrt{1-\theta^2}}$$

and after the $\mathcal{B}$ operators are applied:

$$\|\mathcal{B}(A^\Delta) - \mathcal{B}(A^{CD})\|_S = \sqrt{Var(z_{i,t} - z_{i,t})} + \sqrt{Var(p^\Delta_t - p^{CD}_t)}$$

$$= \sqrt{Var\left(\beta\alpha\xi B^\Delta u_t - \beta B^{CD} u_t + o(\Delta)\right)} = \sqrt{Var\left(\alpha\Delta B^\Delta u_t + \beta\left(B^\Delta u_t - B^{CD} u_t\right) + o(\Delta)\right)}$$

Does this deviation increase the signal norm? The limit as $\Delta \to 0$ is

$$\lim_{\Delta\to 0} \frac{\|\mathcal{B}(A^\Delta) - \mathcal{B}(A^{CD})\|_S}{\|A^\Delta - A^{CD}\|_S} = \lim_{\Delta\to 0} \frac{\sqrt{Var\left(\alpha B^\Delta u_t + \frac{\beta}{\Delta}\left(B^\Delta u_t - B^{CD} u_t\right)\right)}}{\sqrt{1-\theta^2}} \tag{77}$$

This limit can be calculated analytically, but gets complicated quickly, so I will demonstrate numerically that it must be > 1. However, it is worth showing analytically that the $\frac{\beta}{\Delta}\left(B^\Delta u_t - B^{CD} u_t\right)$ term neither diverges as $\Delta \to 0$ nor goes to zero as $\beta \to 0$:

$$\frac{\beta}{\Delta}\left(B^\Delta - B^{CD}\right) = \frac{\beta}{\Delta}\left(\frac{\xi + L}{1+\xi L} - \frac{\theta + L}{1+\theta L}\right)$$

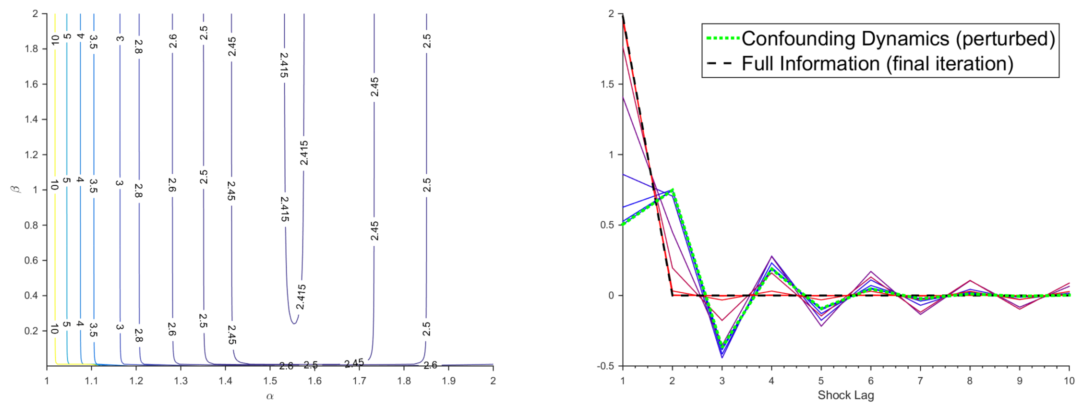
$$= \frac{\beta}{\Delta}\left(\frac{(\xi+L)(1+\theta L) - (\theta+L)(1+\xi L)}{(1+\xi L)(1+\theta L)}\right) = \frac{\beta}{\Delta}\left(\frac{(\xi + L + \xi\theta L + \theta L^2) - (\theta + L + \xi\theta L + \xi L^2)}{(1+\xi L)(1+\theta L)}\right)$$

$$= \frac{\beta}{\Delta}\left(\frac{(\xi-\theta)(1+L^2)}{(1+\xi L)(1+\theta L)}\right) = \frac{(1+L^2)}{(1+\xi L)(1+\theta L)}$$

where the final step uses $\xi = \theta + \frac{\Delta}{\beta}$.

Figure 5 panel (a) demonstrates that the confounding dynamics equilibrium must be signal-unstable, by numerically calculating the proportional change in equation (77) for $\beta \in (0, 1]$ and $\alpha \in [1, 2]$. The minimum deviation in this range is at approximately $\alpha = 1.55$ and $\beta = 1$, and the signal perturbation still more than doubles the signal norm. Indeed, even when considering much larger ranges for the parameters, the proportional increase always appears to be at least 2.4.

Moreover, the confounding dynamics equilibrium is not just locally unstable, failing the technical definition of signal stability. Rather, it is globally numerically unstable! Figure 5 panel (b) demonstrates, plotting a number of IRFs for the price



(a) Proportional Effect of Perturbation

(b) Signal Process Divergence

Figure 5: Instability of the Confounding Dynamics Equilibrium

The left panel plots the proportional signal deviation implied by the perturbation, for each combination of α and β parameters. The right panel shows the instability of the confounding dynamics equilibrium by plotting how the Signal Operator Iteration converges from a small perturbation to the full information equilibrium.

process. I begin by perturbing the confounding dynamics equilibrium to p^{Δ} (dotted green line), and then repeatedly applying the operator $\mathcal{B}$. The perturbation is small ($\Delta = 0.01$), so the signal very slowly begins to diverge but eventually rapidly converges to the full information equilibrium (solid lines with colors shifting from blue to red based on distance to the final equilibrium.) Because the perturbation is small, the initial divergence is hard to see, so I omit many iterations from the plot including the first 20 after the initial perturbation.

Why is this perturbation so explosive even when β is small? When β is near zero, the information feedback is limited, because forecasts are multiplied by a small coefficient when reported as prices. However, when agents make their forecasts, they have to multiply the observed prices by a large $1/\beta$ coefficient. So small perturbations in the price signal can have a large effect on forecasts.

Why was this particular perturbation explosive? Because information feedback regularity is satisfied, perturbations that are spanned by the equilibrium forecast error process w_t cannot have an explosive effect on signals. So if a perturbation can disproportionately move the implied endogenous signal, it should have a large component that is orthogonal to the history of w_t's. This is why I chose the perturbation $\frac{\Delta}{1+\theta L}$; it is orthogonal to the Blaschke factor $L^k \frac{\theta+L}{1+\theta L}$ for all powers $k \geq 1$.

E Time Series in ℓ^2

This Appendix describes how to represent a time series in the Hilbert space ℓ^2.

E.1 Time Series as Vectors

ℓ^2 is the Hilbert space of square-summable infinite sequences. This space is useful for representing time series, and provides an intermediate step towards representing time series as operators.

Consider a stationary time series of the form $x_t = X(L)\varepsilon_t = \sum_{j=0}^{\infty} X_j L^j \varepsilon_t$. If ε_t is scalar-valued, the vector representation of this time series is

$$\vec{x} \equiv \begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \end{pmatrix}$$

Every vector in ℓ^2 maps to a stationary time series in this way. The norm of the vector is its standard deviation.

If ε_t is matrix valued, $\vec{x}$ is a block vector. But this maps back to ℓ^2 by block-vectorizing.

One reason a vector representation is helpful is that a lag operator polynomial of the time series is just a block Toeplitz operator times the vector. For example, if $y_t = A(L)x_t$, then

$$\vec{y} = \begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} A_0 & A_{-1} & A_{-2} & A_{-3} & \ldots \\ A_1 & A_0 & A_{-1} & A_{-2} & \ldots \\ A_2 & A_1 & A_0 & A_{-1} & \ldots \\ A_3 & A_2 & A_1 & A_0 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ \vdots \end{pmatrix}$$

The identically distributed time series x_t has an autocovariance function γ_j where j indicates the order of the autocovariance (i.e. γ_0 is the variance, γ_1 is the first autocovariance, and so forth.) In the ℓ^2 vector representation, the jth autocovariance is the inner product:

$$= \langle \vec{x}, L^j \vec{x} \rangle$$

This generalizes to the matrix-valued case by equation (12).

Why is it useful to represent signals as operators instead of just vectors? One reason is that sometimes the signals need to be right-multiplied, not just left-multiplied. For example, this occurs when applying different white noise processes to a Wold representation, or when aggregating signals across islands by P_G.

E.2 Time Series as Toeplitz Operators

The Toeplitz representation collects the time series $\vec{x}$ into a block Toeplitz operator:

$$\begin{pmatrix} X_0 & 0 & 0 & 0 & ... \\ X_1 & X_0 & 0 & 0 & ... \\ X_2 & X_1 & X_0 & 0 & ... \\ X_3 & X_2 & X_1 & X_0 & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

With this representation, it can be left- or right-multiplied by other operators. If they are causal, the operator is lower triangular, and the output will be causal. Revisiting the example $y_t = A(L)x_t$, the first column of the Toeplitz representation will always be $\vec{y}$. But if $A(L)$ is causal, then the representation is:

$$\begin{pmatrix} Y_0 & 0 & 0 & 0 & ... \\ Y_1 & Y_0 & 0 & 0 & ... \\ Y_2 & Y_1 & Y_0 & 0 & ... \\ Y_3 & Y_2 & Y_1 & Y_0 & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} A_0 & 0 & 0 & 0 & ... \\ A_1 & A_0 & 0 & 0 & ... \\ A_2 & A_1 & A_0 & 0 & ... \\ A_3 & A_2 & A_1 & A_0 & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \begin{pmatrix} X_0 & 0 & 0 & 0 & ... \\ X_1 & X_0 & 0 & 0 & ... \\ X_2 & X_1 & X_0 & 0 & ... \\ X_3 & X_2 & X_1 & X_0 & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

For a concrete example of a Toeplitz representation, consider the VAR(1) process $y_t = By_{t-1} + \varepsilon_t$. The lag operator polynomial for this process is

$$y_t = Y(L)\varepsilon_t = \sum_{j=0}^{\infty} B^j L^j \varepsilon_t$$

which, per equation (19), has block Toeplitz representation

$$[VAR(1)]: \qquad \mathbf{Y} = \begin{pmatrix} I & 0 & 0 & 0 & ... \\ B & I & 0 & 0 & ... \\ B^2 & B & I & 0 & ... \\ B^3 & B^2 & B & I & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

E.3 Connection to Frequency Domain Representations

Consider a time series

$$y_t = Y(L)\varepsilon_t = \sum_{j=0}^{\infty} Y_j L^j \varepsilon_t$$

where ε_t is unit variance white noise. Sometimes, this time-series is analyzed in the frequency domain by defining the "z-transform" $Y(z)$:

$$Y(z) \equiv \sum_{j=0}^{\infty} Y_j z^j \qquad z \in \mathbb{D}$$

such that $Y(z)$ is an analytic function on the unit disc $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. This is the approach taken in Han, Tan, and Wu (2022).

How does the Toeplitz operator $T(Y)$ relate to the analytic function $Y(z)$? The matrix representation is

$$T(Y) = \begin{pmatrix} Y_0 & 0 & 0 & 0 & ... \\ Y_1 & Y_0 & 0 & 0 & ... \\ Y_2 & Y_1 & Y_0 & 0 & ... \\ Y_3 & Y_2 & Y_1 & Y_0 & ... \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The entries Y_j are the *Fourier coefficients* of the function $Y(z)$. The function $Y(z)$ is called the *symbol* of the Toeplitz operator $T(Y)$. When the symbol is analytic, many operations on the symbols and Toeplitz operators are analogous. For example, given analytic functions $X(z)$ and $Y(z)$:

1. Linear transformations satisfy $aT(X) + bT(Y) = T(aX + bY)$
2. Multiplication satisfies $T(X)T(Y) = T(XY)$
3. Inversion satisfies $T(Y^{-1}) = T(Y)^{-1}$

Lastly, a symbol is analytic if its Toeplitz operator is block triangular. This is why analytic functions are useful for representing causal time series.[34]

[34] Böttcher and Silbermann (2013) is the reference for this section and is a valuable resource regarding Toeplitz operators.