

CONFOUNDING DYNAMICS*

Giacomo Rondina[†] Todd B. Walker[‡]

January 2021

ABSTRACT

In the context of a dynamic model with incomplete information, we isolate a novel mechanism of shock propagation. We term the mechanism *confounding dynamics* because it arises from agents' optimal signal extraction efforts on variables whose dynamics—as opposed to superimposed noise—prevents full revelation of information. Employing methods in the space of analytic functions, we are able to obtain analytical characterizations of the equilibria that generalize the celebrated Hansen-Sargent optimal prediction formula. Our main theorem establishes conditions under which confounding dynamics emerge in equilibrium in general settings. We apply our results to a canonical one-sector real business cycle model with dispersed information. In that setting, confounding dynamics is shown to amplify the propagation of a productivity shock, producing hump-shaped impulse response functions.

Keywords: Dispersed Information, Confounding Dynamics, Rational Expectations

*We would like to thank Manuel Amador, Marios Angeletos, Ryan Chahrour, Tim Cogley, Thorsten Drautzburg, Jesús Fernández-Villaverde, Zhen Huo, Jennifer La'O, Eric Leeper, Thomas Mertens, Kristoffer Nimark, Gary Ramey, Tom Sargent, Martin Schneider, Dimitri Vayanos, Venky Venkateswaran, Pierre-Olivier Weill, Mirko Wiederholt, Chuck Whiteman and seminar participants at UC–Berkeley, UC–Santa Cruz, UC–San Diego, New York University, University of Pennsylvania, University of Texas–Austin, Stanford University, St. Louis Federal Reserve, Richmond Federal Reserve, Carnegie Mellon University, the 2017 Society of Economic Dynamics Annual Meeting in Edinburgh, and the 2017 Society for Computational Economics International Conference in New York. We would also like to thank an associate editor and two anonymous referees for helpful comments. We also acknowledge financial support from the National Science Foundation under grant number SES–0962221.

[†]University of California – San Diego, grondina@ucsd.edu

[‡]Indiana University, walkertb@indiana.edu

1 INTRODUCTION

Modeling and seeking to understand economic fluctuations is one of the cornerstones of modern economics. The role of incomplete information in this endeavor was acknowledged very early on by Pigou (1929) and Keynes (1936). Their ideas were first formalized in a rational expectations setting by Lucas (1972, 1975), King (1982) and Townsend (1983b). The underlying theme that ties these papers together is that unresolved uncertainty—in and of itself—can be a source of fluctuation in the economy. This idea has seen a resurgence. Dynamic models with dispersed information are becoming increasingly prominent in several literatures such as asset pricing, optimal policy communication, international finance, and business cycles.¹ Our paper contributes to this literature by introducing a novel mechanism of shock propagation, which we call confounding dynamics, and does so in a manner that permits tractability.

Confounding dynamics arise from optimal prediction (i.e. rational expectations) in which past realizations of economic shocks prevent full revelation of information today, even when an arbitrarily large amount of data is available. Ensuring confounding dynamics emerge in equilibrium amounts to deriving non-invertibility restrictions on the equilibrium system of equations. If this system is non-invertible in current and past observations, agents will never fully unravel the contemporaneous economic shock. Our primary example of Section 5, which is based on the real business cycle model of Lucas (1975), shows that non-invertibility of the exogenous process is not a necessary condition for confounding dynamics. The model’s cross-equation restrictions endogenously generate non-invertible representations, even when the exogenous process is always invertible. Confounding dynamics can also persist when the number of observables is equal to the number of shocks and therefore, our approach does not rely on the need to overrun the agent’s information set with exogenous noise.

We articulate the idea of confounding dynamics in three steps. First, Section 2 derives an optimal prediction formula under confounding dynamics that extends the celebrated Hansen-Sargent formula, and makes an explicit connection to these dynamics. Subsequently, we demonstrate that this behavior carries over to a generic rational expectations model with dispersed information. Our main theorem contains two equations—one that characterizes the dynamic properties of the equilibrium when confounding dynamics are present and one that derives restrictions that guarantee confounding dynamics are preserved in equilibrium. Finally, we provide economic intuition by introducing confounding dynamics into a standard Real Business Cycle model. This application showcases the central insight coming from our main theorem and the defining property of confounding dynamics. The insight is that permitting information to arise endogenously in the

¹The literature is too voluminous to cite every worthy paper. Recent examples include: Woodford (2003a), Pearlman and Sargent (2005), Allen, Morris, and Shin (2006), Bacchetta and van Wincoop (2006), Hellwig (2006), Adam (2007), Gregoir and Weill (2007), Angeletos and Pavan (2007), Kasa, Walker, and Whiteman (2014), Lorenzoni (2009), Rondina (2009), Angeletos and La’O (2009), Angeletos and La’O (2013), Hellwig and Venkateswaran (2009), Graham and Wright (2010), Nimark (2010), Hassan and Mertens (2011), Benhabib, Wang, and Wen (2015), Huo and Takayama (2016) and Angeletos and Lian (2016).

context of a model opens the door to an equilibrium that is usually overlooked when information is exogenously provided to agents. Our analytical representation allows us to carefully show how confounding dynamics interacts with crucial parameters of the model. For example, as the elasticity of substitution increases, endogenous variables become more informative and it is more difficult to maintain confounding dynamics in equilibrium. The defining property of confounding dynamics is an impulse response function that is amplified and more persistent relative to the full information equilibrium. There are two possible shapes of an impulse response to a fundamental shock under confounding dynamics: [i.] fluctuations around the full information counterpart that display the “waves of optimism and pessimism” of Pigou (1929); and [ii.] an amplified impulse response function that is hump-shaped. We discuss both scenarios in the context of exogenous signal extraction in Section 2. Section 5 focuses on the latter type of impulse response and argues that confounding dynamics—without additional frictions—can provide the internal propagation necessary to match important moments of the data along the lines discussed in Cogley and Nason (1995).

We solve and analyze the rational expectations equilibrium in the space of analytic functions. This approach has several advantages vis-a-vis standard time-domain methods. For example, as emphasized in Townsend (1983a), equilibria are sought in generic functional spaces spanned by linear combinations of shocks, which allows one to avoid explicitly modeling higher-order belief dynamics. Moreover, the matrix Ricatti equation typical of Kalman filtering is replaced by a more transparent spectral factorization problem. This allows us to solve and analyze the equilibrium in closed form. We are not the first to advocate such an approach. Others, such as Futia (1981), Townsend (1983a), Taub (1989), Kasa (2000), Walker (2007), Rondina (2009), Bernhardt, Seiler, and Taub (2010), Kasa, Walker, and Whiteman (2014), and Huo and Takayama (2016) have used similar techniques to solve dynamic rational expectation models with incomplete information. We contribute to this literature by deriving analytical representations (e.g., generalized Hansen-Sargent formulas) and by providing a systematic treatment of equilibrium conditions in models with *dispersed* information that display confounding dynamics. Futia (1981) and Townsend (1983a) were the first to advocate for the use of analytic functions to solve dynamic rational expectations models with heterogeneous information. Many of the mathematical antecedents of this paper can be found there and in Whiteman (1983). Taub (1989) demonstrates how the algebra associated with dynamic signal extraction (i.e., spectral factorization) is simplified through the analytic function approach. We take advantage of these formulas to completely characterize existence and uniqueness of equilibria in dispersed informational setups. Bernhardt, Seiler, and Taub (2010) and Kasa, Walker, and Whiteman (2014) do not examine models with dispersed information, but show how these methods can be used to help resolve asset pricing anomalies.

2 PREDICTION WITH CONFOUNDING DYNAMICS

To study our primary mechanism, we present a simple version of the prediction problem that operates at the heart of the rational expectations equilibria with confounding dynamics. For the reader unfamiliar with frequency domain methods we provide a primer in Appendix C.

Consider the univariate process specified as

$$s_t = -\lambda\varepsilon_t + \varepsilon_{t-1} = (L - \lambda)\varepsilon_t, \quad (1)$$

where ε_t is a mean-zero, normally distributed variable with variance σ_ε^2 . Suppose that the prediction problem is to compute the mean-squared error minimizing prediction for ε_t given that s^t is observed. To fix ideas and foreshadow results, imagine that ε_t is the time- t unobserved innovation in aggregate productivity in the economy, while s_t is the observed market rental rate of physical capital. The prediction problem asks for an estimate of the current productivity innovation using the history of the market rental rate.

To solve the problem, we need to consider two possible cases. If $|\lambda| \geq 1$, the process is deemed fundamental for ε_t using the terminology of Rozanov (1967), which means that the stochastic process (1) is invertible in current and past observables; therefore there exists a linear combination of current and past s_t 's that allows the exact recovery of ε_t . Defining the lag operator $Lx_t = x_{t-1}$, one can easily verify that with $|\lambda| \geq 1$, $L - \lambda$ is an invertible operator, and the optimal prediction corresponds to

$$\mathcal{P}(\varepsilon_t | s^t) = \frac{s_t}{L - \lambda} = -\frac{1}{\lambda}(s_t + \lambda^{-1}s_{t-1} + \lambda^{-2}s_{t-2} + \lambda^{-3}s_{t-3} + \dots) = \varepsilon_t, \quad (2)$$

which verifies that the history of s^t contains all the information needed to perfectly know ε_t .

Consider now the case of $|\lambda| < 1$. Clearly, the prediction formula (2) is no longer well defined as the coefficients diverge. In this simple environment, Rozanov (1967) shows that the appropriate factorization requires flipping the root λ outside of the unit circle through the use of a Blaschke factor, which we denote as $\mathcal{B}(L) = (1 - \lambda L)/(L - \lambda)$.² Applying the Blaschke factor results in the optimal prediction,

$$\mathcal{P}(\varepsilon_t | s^t) = -\frac{\lambda}{1 - \lambda L}s_t = -\lambda(s_t + \lambda s_{t-1} + \lambda^2 s_{t-2} + \lambda^3 s_{t-3} + \dots) = -\lambda \left(\frac{L - \lambda}{1 - \lambda L} \right) \varepsilon_t, \quad (3)$$

Note that the mean squared forecast error of $(1 - \lambda^2)\sigma_\varepsilon^2 > 0$, demonstrating that as $|\lambda|$ approaches one from below there is exact recovery of ε_t .

When the process is non-invertible, (3) shows that the history of current and past s_t 's reveals a particular linear combination of ε_t 's. Expanding this last term yields

$$\mathcal{P}(\varepsilon_t | s^t) = \underbrace{\lambda^2 \varepsilon_t}_{\text{information}} - \underbrace{(1 - \lambda^2)[\lambda \varepsilon_{t-1} + \lambda^2 \varepsilon_{t-2} + \lambda^3 \varepsilon_{t-3} + \dots]}_{\text{noise from confounding dynamics}}. \quad (4)$$

²Specifically, the Blaschke factor flips the zero from inside the unit circle to outside the unit circle via the transformation

$$(L - \lambda) \left(\frac{1 - \lambda L}{L - \lambda} \right) \left(\frac{L - \lambda}{1 - \lambda L} \right) \varepsilon_t$$

Note that $\mathcal{B}(L)\mathcal{B}(L)^{-1} = 1$, and therefore, the Blaschke factor does not alter the covariance generating function of the time series.

Thus, the noise resulting from confounding dynamics takes an unusual form as it consists of a linear combination of past realizations of ε_t . Expression (4) suggests that the process (1) is informationally equivalent to a noisy signal about ε_t , where the noise is the linear combination of past shocks (in the bracketed term), and the signal-to-noise ratio is measured by λ^2 . A λ closer to zero results in less information and more noise but, at the same time, it also makes past shocks less persistent. As $\lambda \rightarrow 0$, there is no information in s_t about ε_t and the optimal prediction is 0, the unconditional average. As long as $|\lambda| \in (-1, 1)$, the value of ε_t will *never* be learned and in this sense, the *history* of the fundamental shock acts as a noise shock but (as shown below) has non-standard properties. This is the defining property of confounding dynamics. The shocks are perfectly correlated and no super-imposed noise process is necessary to keep full revelation of information from occurring. An infinite history of past shocks is not sufficient because the dynamic history of the shock confounds agents into making forecast errors that would be persistent under the standard full-information rational expectations case.

2.1 ECONOMIC INTERPRETATION We now provide some economic intuition as it relates to our signal extraction problem, noting that additional intuition is found in Section 5, where we embed this learning mechanism in a real business cycle model.

Comparing representation (1), which we repeat here for convenience, $s_t = (L - \lambda)\varepsilon_t$, to the fundamental representation used to form the optimal prediction (2), $s_t = (1 - \lambda L)\tilde{\varepsilon}_t$ where $\tilde{\varepsilon}_t = \mathcal{B}(L)^{-1}\varepsilon_t$, we see that information is discounted differently. Under full information (assuming agents observe the underlying shocks directly), last period's shock would be discounted more heavily relative to the contemporaneous shock, recall $|\lambda| < 1$. This discounting is exactly reversed when agents have confounding dynamics (assuming agents only observe s^t) with the contemporaneous shock receiving the more significant discount. Therefore, innovations entering the agents' information sets will be discounted differently from the full-information case when confounding dynamics is operational. The extent of the difference in discounting is dictated entirely by the parameter λ : as λ approaches zero (one), the difference will be large (small).

An alternative interpretation comes from noting that confounding dynamics nests the sticky information setup of Mankiw and Reis (2002). When $\lambda = 0$, innovations are observed by agents with a one-period lag, in accordance with sticky information. One might argue that this assumption is too strong in that agents may not ignore *all* information with a one-period lag. Our representation allows for a more continuous interpretation. As $|\lambda|$ approaches one from below starting from zero, agents become more informed. For $|\lambda| \geq 1$, all information is revealed. In principle, one could estimate this parameter using standard methods in a DSGE model. The estimate of λ would then determine the optimal amount of “stickiness” as dictated by data. Several papers argue that sticky information is a natural setup because it can reconcile macro price rigidity with micro price flexibility [Klenow and Willis (2007)] and survey expectations of inflation [Coibion and Gorodnichenko (2012)]. Our approach suggests there is even more flexibility along this dimension.

Finally, we note that the econometrics literature has seen a renewed interest in identification of vector auto-regressions (VAR) in the presence of non-invertibilities [see, Canova and Sahneh (2017)].

One argument in favor of confounding dynamics is that if econometricians using sophisticated techniques have trouble cleanly identifying shocks to the macroeconomy, agents will most likely suffer from similar identification problems, implying non-invertibilities are more likely than not. In this instance, theory can help with measurement because we, as modelers, can cleanly identify ε_t from $\tilde{\varepsilon}_t$, and can then ask how the economy responds to the structural innovation, ε_t , when agents have incomplete information.

2.2 CONNECTION TO STANDARD SIGNAL EXTRACTION To make the connection to the standard signal extraction problem more explicit, suppose that agents observe an infinite history of the signal

$$x_t = \varepsilon_t + \eta_t, \tag{5}$$

where $\eta_t \stackrel{iid}{\sim} N(0, \sigma_\eta^2)$. The optimal prediction is well known and given by $\mathcal{P}(\varepsilon_t|x^t) = \tau x_t$, where τ is the relative weight given to the signal, $\tau = \sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + \sigma_\eta^2)$. It can be shown³ that the information content of (1) with $|\lambda| < 1$ is equivalent to (5), where equivalence is defined as equality of variance of the forecast error conditioned on the infinite history of the observed signal, i.e.

$$\mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t|s^t))^2 \right] = \mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t|x^t))^2 \right],$$

when

$$\lambda^2 = \tau. \tag{6}$$

Notice that when the signal-to-noise ratio increases (decreases), this corresponds to a higher (lower) absolute value of λ . In the limit, as $\sigma_\eta^2 \rightarrow 0$, then $\lambda^2 \rightarrow 1$, which ensures exact recovery of the state in both cases.

While the informational content can be made identical, the dynamics of the two signal extraction problems are very different. To visualize this, we report the impulse response function for the prediction equations that contain confounding dynamics (4) and for the standard signal extraction problem (5) to a one time, one unit increase in ε_t in Figure 1. We do this for both a low and high value of λ^2 (resp. τ).⁴

Figure A reports a negative value for the non-invertible root λ . Here the impulse response to (4) under-predicts the actual innovation on impact (which is one), with a smaller value of λ under-predicting more significantly. This is due to the first term on the RHS of (4). The same is true for the standard signal extraction formulation (dashed lines). Agents weigh the initial innovation by the signal-to-noise ratio $\tau < 1$ and therefore under-predict on impact. This is where the similarities end. With confounding dynamics, periods two through six show waves of over- and under-prediction relative to the actual realization and relative to the standard signal extraction problem. As discussed above, the current and past innovations will persistently affect the prediction function several periods beyond impact. This defining characteristic of confounding

³See Online Appendix B.2 for a proof.

⁴For aesthetic reasons, the impulse responses are slightly smoothed at turning points.

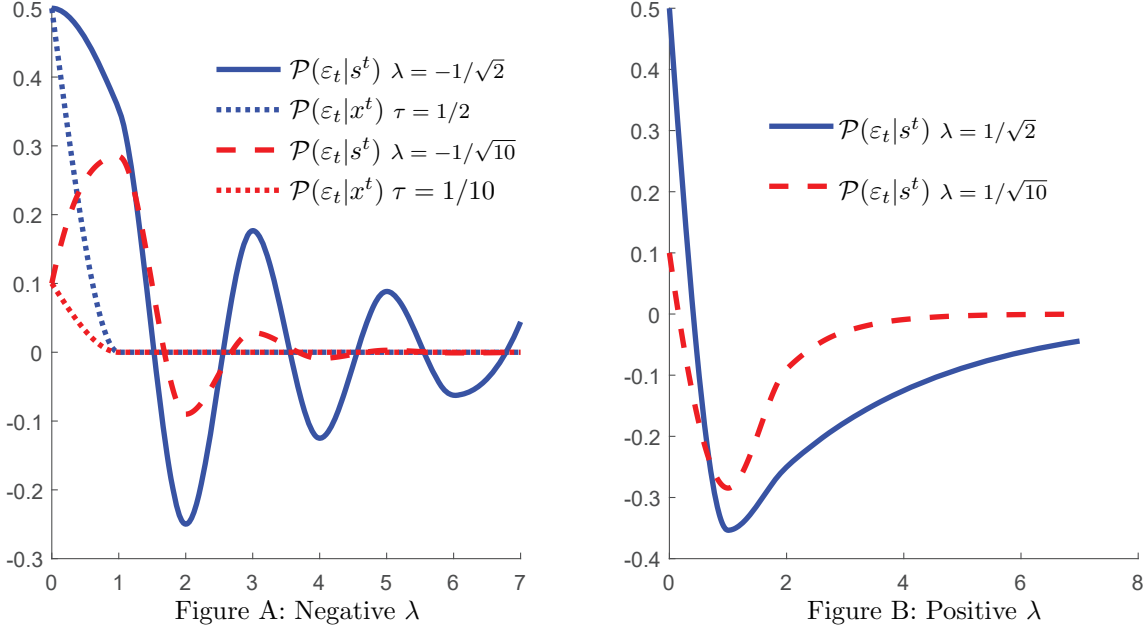


Figure 1: Panel A: Impulse Responses of x_t and s_t to a one unit change in ε_t for signal-to-noise ratios of $\tau = 1/2$, $\lambda = -1/\sqrt{2}$ (dotted, solid blue) and $\tau = 1/10$, $\lambda = -1/\sqrt{10}$ (dotted, dashed). Panel B: Impulse response of x_t for $\lambda = 1/\sqrt{2}$ (solid) and $\lambda = 1/\sqrt{10}$ (dashed).

dynamics leads to the waves of over- and under-reaction. This is in contrast to the full information case and standard signal-extraction case where the impulse response is zero after impact. As already pointed out, the smaller the λ , the larger the noise term in (4), but the less persistent the over- and under-prediction. Thus optimal signal extraction with confounding dynamics generates fluctuations where the full-information and exogenously imposed noise counterparts generate none. Figure B shows that the under- and over-reaction is not the only form of the impulse response under confounding dynamics. A positive value for λ generates an (inverse) hump-shaped impulse response.⁵ Again, this can be seen from (4): the under-reaction on impact is the same independent of sign due to the λ^2 term; a positive value for λ implies that the elements of the noise term of (4) all enter with the same sign, causing the impulse to return gradually from below. The larger the value of λ , the more the impulse overshoots. Therefore in either case, confounding dynamics adds persistence to the impulse where traditional signal extraction would not.

3 MODEL, INFORMATION, AND EQUILIBRIUM

We now model confounding dynamics in a generic rational expectations formulation that permits many interpretations (e.g., monetary model, asset pricing model, etc.). We do this via dispersed information, which introduces well-known difficulties. We lay out a solution strategy and compare that strategy to alternative methodologies.

⁵In a different setting, Acharya, Benhabib, and Huo (2017) show that the combination of sentiment shocks and non-invertibilities can generate hump-shaped impulse response functions as well.

3.1 MODEL We consider models that are populated by a continuum of agents indexed by $i \in [0, 1]$. Let $\mu(i)$ be the density of agent i characterized by the information set at time t , denoted by Ω_{it} . We are interested in the class of models in which the individual optimal choice can be represented by the dynamic expectational difference equation,

$$\phi \mathbb{E}[X_{it+1} | \Omega_{it}] = \psi(L)X_{it}, \quad (7)$$

where

$$X_{it} \equiv \begin{pmatrix} x_{it} & y_t & \theta_{it} \end{pmatrix}^\top \quad (8)$$

Here $\phi \equiv [\phi_x \ \phi_y \ \phi_\theta]$, is a vector of coefficients, and $\psi(L) \equiv [\psi_x(L) \ \psi_y(L) \ \psi_\theta(L)]$, is a vector of square-summable lag polynomials in non-negative powers of L . x_{it} is the choice variable under the control of the individual agent i ; y_t is an endogenous aggregate variable that agents take as given, and θ_{it} is an exogenous stochastic process specified as the sum of an aggregate component θ_t and an i.i.d. individual component v_{it} . Formally

$$\theta_{it} = \theta_t + v_{it}, \quad \text{where } \theta_t = A(L)\varepsilon_t, \quad (9)$$

with $v_{it} \sim \mathcal{N}(0, \sigma_v)$, $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$, and $A(L)$ is a square-summable polynomial in non-negative powers of L . Our main theorem will deliver the restrictions on parameters needed to ensure the equilibrium system of equations is non-invertible in current and past observations; i.e., that confounding dynamics obtains in equilibrium. To close the model we need to specify a relationship between the distribution of x_{it} across agents, and the aggregate y_t . We thus posit that

$$\gamma(L) \int_0^1 X_{it} \mu(i) di = 0, \quad (10)$$

where $\gamma(L) \equiv [\gamma_x(L) \ \gamma_y(L) \ \gamma_\theta(L)]$, is a vector of square-summable finite-degree lag polynomials in non-negative powers of L , and we assume $\gamma_x(L) \neq 0$.⁶ As we proceed with the analysis it will be useful to think of equation (7) as representing a demand (or supply) schedule for agent i , and (10) as the relevant market clearing condition. However, the specific form depends on the particular application at hand.

The expectational difference equation (7) is a dispersed information version of the system originally considered by Blanchard and Kahn (1980), and subsequently studied by Uhlig (1999), Klein (2000) and Sims (2002), among others. Dispersed information implies that individual expectations are heterogeneous, which implies that the aggregation in (10) will result in taking an average of expectations. In particular, model (7)-(10) can accommodate both average expectations of aggregate variables and average expectations of individual variables.

⁶We make this assumption in order to keep the connection between (7) and (10) non-trivial. Allowing for $\gamma_x(L) = 0$, would imply that y_t is directly determined by the process θ_t , and, as a consequence, it would enter (7) as an exogenous variable, essentially duplicating the role of θ_t in that equation.

3.2 INFORMATION In our dispersed information setup, we assume that the information set Ω_{it} of an arbitrary agent i at time t consists of the smallest closed subspace generated by the history of the random variable $\theta_i^t \equiv \{\theta_{it}, \theta_{it-1}, \dots\}$, and the history of the aggregate variable $y^t = \{y_t, y_{t-1}, \dots\}$. Specifically, $\Omega_{it}^i = \theta_i^t \vee y^t$, where the operator \vee denotes the span (i.e., the smallest closed subspace which contains the subspaces) generated by the sequences θ_i^t and y^t . This notation simply suggests that expectations will be taken optimally; i.e., they will be consistent with the prediction formulas discussed in Section 2. In a multivariate moving-average setting, the invertible representation achieved via canonical factorization is the smallest closed subspace containing the observables, θ_i^t and y^t (see Hoffman (1962)).

Given (7), x_{it} will be a function of the history of idiosyncratic innovations, v_{it} , and the aggregate innovations, ε_t , namely

$$x_{it} = X(L)\varepsilon_t + V(L)v_{it}. \quad (11)$$

In addition, aggregation implies that y_t is only a function of aggregate innovations, so that

$$y_t = Y(L)\varepsilon_t. \quad (12)$$

The signal structure can be thus represented as

$$\begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix} = \Gamma(L) \begin{pmatrix} \sigma_\varepsilon^{-1} \varepsilon_t \\ \sigma_v^{-1} v_{it} \end{pmatrix}, \quad \Gamma(L) = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ Y(L)\sigma_\varepsilon & 0 \end{bmatrix}. \quad (13)$$

We point out that our information set is in line with the typical information set assumed in the dispersed information rational expectations literature: we provide agents with both an *exogenous* signal about the aggregate unobserved state (θ_{it}), and an *endogenous* signal that is determined in equilibrium (y_t). The analytical convenience of the signal structure (13), for our purposes, is that the invertibility of the matrix $\Gamma(L)$ hinges only upon the zeros of $Y(L)$. At the same time, the structure imposes analytical discipline that is uncommon in the literature: the endogenous signal y_t can reveal perfectly the underlying state, under the appropriate parametrization of model (7)-(10). Thus, we aim at establishing both the degree to which information remains incomplete in equilibrium, along with the more standard existence and uniqueness conditions.

3.3 EXAMPLES We pause briefly here to note that our general setup can handle a wide variety of models. Appendix B.7 carefully walks readers through four such examples: an RBC model, the asset pricing model of Singleton (1987), a model with Calvo pricing and a New Keynesian Phillips Curve, and the classical monetary models of inflation of Cagan (1956). Of course, this list is not exhaustive but there are two common characteristics in all of the examples: [i.] shocks are Gaussian and [ii.] the model can be written in a linear form. As with nearly all papers in this literature, our analysis relies on linear projections being consistent with optimal conditional expectations, which necessitates [i] and [ii].

3.4 EQUILIBRIUM DEFINITION Uncertainty is assumed to be driven by Gaussian innovations, which, together with linearity, implies that conditional expectations are computed as optimal linear projections. We thus have

$$\mathbb{E}(X_{it+1}|\Omega_{it}) = \mathcal{P}[X_{it+1}|\Omega_{it}], \quad (14)$$

and can apply the Wiener-Kolmogorov prediction formula (see Appendix C) to compute conditional expectations. We are now ready to define a Rational Expectations Equilibrium for model (7)-(10).

Definition REE. *A Rational Expectations Equilibrium (REE) is a stochastic process for $\{X_{it}, i \in [0, 1]\}$ and a stochastic process for the information sets $\{\Omega_{it}, i \in [0, 1]\}$ such that: (i) each agent i , given her information set, forms expectations according to (14); (ii) $\{X_{it}, i \in [0, 1]\}$ satisfies conditions (7)-(10).*

The REE can be summarized by two statements: (a) given a distribution of information sets, there exists a market clearing distribution $\{X_{it}, i \in [0, 1]\}$ determined by each agent i 's optimal prediction conditional on the information sets; (b) given a distribution $\{X_{it}, i \in [0, 1]\}$, there exists a distribution of information sets that provides the basis for optimal prediction. Both statements (a) and (b) must be satisfied by the same distribution $\{X_{it}, i \in [0, 1]\}$ and the same distribution of information sets *simultaneously* in order to satisfy the requirements of a REE. This dual fixed point condition is standard in rational expectations with potentially heterogeneously informed agents and when endogenous variables convey information [see, Radner (1979) as an early example].

3.5 WEIGHTED SUM OF EXPECTATIONS Before discussing our solution methodology, we give a brief overview of the typical approach to solve model (7)-(10), which consists of two steps. The first step is to iteratively substitute the endogenous variables x_{it+j} and y_{t+j} forward by leading (7) j periods forward and aggregating over agents. The end result is expressions for x_{it} and y_t , that are a function of expectations of the exogenous variable θ_t at all future horizons. The second step is then to compute those expectations, which is non-trivial due to the fact that the law of iterated expectations may not be operational. Most of the work that uses this approach rely on numerics to calculate these expectations.⁷

Consider the expression for $\phi_x \mathbb{E}_{it}(x_{it+1})$. Through forward substitution, this expression contains the term $\phi_x \mathbb{E}_{it+1}(x_{it+2})$, which in turn contains θ_{t+2} . It follows that the law of iterated expectations (LIE) applies in this context so that $\phi_x^2 \mathbb{E}_{it} \mathbb{E}_{it+1}(\theta_{t+2}) = \phi_x^2 \mathbb{E}_{it}(\theta_{t+2})$, and aggregation implies $\phi_x^2 \bar{\mathbb{E}}_t(\theta_{t+2})$ for $j = 2$. Intuitively, in each round of the iterative substitutions there are terms where agent i is taking expectations of both her own future expectations and of future average expectations. The law of iterated expectations applies to the former, so that the order of expectations is reduced, but not to the latter.⁸ It should be evident at this point that the second

⁷Nimark (2010), and Melosi (2016) are recent examples of sophisticated numerical methods to characterize equilibria with dispersed information.

⁸Mechanically, whether LIE applies or not at each iteration depends on the position of ϕ_x in the coefficients of the polynomial $(\phi_x + \phi_y)^j$, i.e. on the set of permutations of size j of ϕ_y and ϕ_x with repetition. For instance, for the case of $j = 2$, the set of terms that multiply ψ_θ are $(\phi_y^2 + \phi_y \phi_x) \bar{\mathbb{E}}_t \bar{\mathbb{E}}_{t+1}(\theta_{t+2}) + (\phi_x \phi_y + \phi_x^2) \bar{\mathbb{E}}_t(\theta_{t+2})$.

step required by the canonical approach—computing closed form solutions for the expectations of arbitrary order—is a daunting task under dispersed information (for more details on this, see Appendix B.5). As already remarked and discussed thoroughly in the next section, we approach the solution from a different angle.

3.6 SOLUTION METHODOLOGY Our aim is to characterize a REE equilibrium for model (7)-(10) with confounding dynamics. The critical requirement for confounding dynamics to emerge is that the information matrix $\Gamma(L)$, (13), must be non-invertible at a $\lambda \in (-1, 1)$. However, there is no guarantee that this condition will hold. Consistent with the intuition of Townsend (1983a), our approach is to formulate a guess for the endogenous variables that follows a generic polynomial in the underlying shocks, and then derive conditions on the *exogenous* parameters that yield non-invertibility in equilibrium.

Our main theorem (Theorem 1) and corollary in Section 5 restricts attention to functional forms with exactly one λ inside the unit circle. The solution procedure described below is consistent with this restriction. However, equilibrium conjectures of functional forms with multiple λ 's inside the unit circle can be entertained within the procedure described below with appropriate modifications. Appendix B.4 shows how to solve the exogenous signal extraction problem with multiple roots inside the unit circle, which provides a road map for how to modify Steps 1-4 below to solve for the rational expectations equilibrium in that case.

The following steps describe our procedure when looking for an equilibrium with confounding dynamics.

1. Specify the guesses for x_{it} and y_t as generic polynomials of underlying shocks

$$x_{it} = X(L)\varepsilon_t + V(L)v_{it}, \quad \text{and} \quad y_t = Y(L)\varepsilon_t. \tag{15}$$

where y_t has confounding dynamics, so that

$$Y(\lambda) = 0, \quad \text{for} \quad \lambda \in (-1, 1). \tag{16}$$

2. Given the signal matrix $\Gamma(L)$, obtain the so-called canonical factorization $\Gamma^*(L)$ under (16) (see Appendix C for a discussion of the canonical factorization).
3. Use $\Gamma^*(L)$ together with the guesses in (15) to obtain the conditional expectations in (7) via the Wiener-Kolmogorov prediction formula.
4. Aggregate over agents according to (10) and use the relationship between $X(L)$ and $Y(L)$ to substitute $X(L)$ with $Y(L)$ in (7). Both the right hand side and the left hand side will now be lag polynomial operators in ε_t and v_{it} , and will thus provide the fixed point conditions for $Y(L)$ and $V(L)$.
5. Derive conditions on exogenous parameters so as to ensure that a determinate stationary

solutions exists, and that there exists a $|\lambda| < 1$, verifying (16). Once $Y(L)$ is solved for, use (10) to recover $X(L)$.

Note that at no point in the solution procedure one needs to worry about higher-order expectations. The so-called “higher-order thinking” that complicates the iterative approach outlined in Section 3.5 is implicit in how the guess (15) combines with the information matrix $\Gamma(L)$ to provide a closed form for the first order expectations in (7). As recognized by Townsend (1983a), by guessing a generic lag polynomial, the higher-order beliefs are built into the guess and we do not have to track these terms explicitly, although higher-order beliefs can be backed out of the solution in closed form. The same solution procedure is followed when we solve for an equilibrium with full information, with the only difference that condition (16) is not imposed, and thus does not have to be verified, and the signal matrix $\Gamma(L)$ corresponds to full information.

4 EQUILIBRIUM WITH CONFOUNDING DYNAMICS

This section establishes the main result of the paper: the existence of a rational expectations equilibrium with confounding dynamics in a dispersed information environment.

4.1 EQUILIBRIUM WITH CONFOUNDING DYNAMICS: MAIN THEOREM In this section we state our main Theorem, which provides conditions under which a REE with Confounding Dynamics exists. As stated in Section 3.2, we specify the information set as

$$\Omega_{it} = \theta_i^t \vee y^t \tag{17}$$

Agents thus observe the entire history of the exogenous process θ_{it} up to time t , together with the history of the aggregate variable y_t . In addition, the model equations (7)-(10) are both common knowledge across agents.

An important building block in the statement of our main theorem is the full information benchmark solution which we denote by, $x_{it} = \mathcal{X}(L)\varepsilon_t + \mathcal{V}(L)v_{it}$, and $y_t = \mathcal{Y}(L)\varepsilon_t$, where $\mathcal{X}(L)$, $\mathcal{V}(L)$ and $\mathcal{Y}(L)$ are square-summable lag polynomial in non-negative powers of L . In the full information solution, each agent i is provided with the entire history of shocks, ε_t and v_{it} , up to time t . The derivation of the full information solution is reported in Appendix A.1. Here we point out that to ensure uniqueness (determinacy) of a full information solution, the characteristic polynomials of the expectational difference equations for $\mathcal{V}(L)$ and $\mathcal{Y}(L)$, which are defined respectively by $\phi_x(L) \equiv \phi_x - \psi_x(L)$, and $\Phi(L) \equiv \phi_x(L) + \phi_y - \psi_y(L)L$, must satisfy a standard regularity condition, which corresponds to our Assumption 1.

Assumption 1: The polynomials $\phi_x(L)$ and $\Phi(L)$ each have exactly one root inside the unit circle.

It is important to note that this assumption does not correspond to a special case, nor is overly restrictive. It amounts to restricting equilibria to stationary processes both within the cross-section and time series dimensions of the model. Requiring $\Phi(L)$ to have one root inside the unit circle is

the standard assumption necessary to yield a unique rational expectations equilibrium (e.g., Sims (2002)) and it immediately implies that $\Phi(L)$ can be factorized as

$$\Phi(L) = (\zeta - L)\tilde{\Phi}(L), \quad (18)$$

where $|\zeta| < 1$, and $\tilde{\Phi}(L)$ has no roots inside the unit circle. If the polynomial had no such roots inside the unit circle, the RE equilibrium would not be unique; and if the polynomial had multiple roots inside the unit circle, no stationary equilibrium would exist. Similarly, requiring $\phi_x(L)$ to have one root inside the unit circle ensures that the cross-sectional distribution is well defined at any point in time. In the equilibrium with confounding dynamics, the expectational difference equations for $V(L)$ and $Y(L)$ contain the same characteristic polynomials $\phi_x(L)$ and $\Phi(L)$ of the full information benchmark, and we thus also impose the regularity conditions of Assumption 1 in Theorem 1.

Recall that the key requirement in solving for an equilibrium with confounding dynamics is that there exists a $\lambda \in (-1, 1)$ such that $Y(\lambda) = 0$. We are interested in finding restrictions on exogenous parameters so that a λ that satisfies such requirement exists. Theorem 1 states our main result.

Theorem 1. *Consider model (7)-(10) with Assumption 1. Let the information sets be specified as in $\Omega_{it} = \theta_i^t \vee y^t$. There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form, $y_t = Y(L)\varepsilon_t$, with*

$$Y(L) = \mathcal{Y}(L) - (1 - \tau(\lambda))(1 - \lambda^2) \frac{\mathcal{A}(\lambda)}{(1 - \lambda L)\tilde{\Phi}(L)}, \quad (19)$$

if there exists a $\lambda \in (-1, 1)$ that solves

$$\mathcal{Y}(\lambda)\tilde{\Phi}(\lambda) = (1 - \tau(\lambda))\mathcal{A}(\lambda), \quad (20)$$

where $\mathcal{Y}(L)$ is the full information solution, $\tau(\lambda) \equiv \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}$, $\mathcal{A}(\lambda)$ is a function of λ that depends only on exogenous parameters, and $Y(L)$ in (19) has a zero inside the unit circle equal to λ .

Proof. See Appendix A.2. □

Theorem 1 provides sufficient conditions for the existence of an equilibrium that belongs to a class in which $Y(L)$ takes a functional form with *exactly one zero* inside the unit circle, that is $Y(L) = (L - \lambda)G(L)$, where $G(L)$ is a stationary lag polynomial with no zeros inside the unit circle. Within the “exactly one zero” class, condition (20) might be satisfied by more than one numerical value for λ . Each value corresponds to a legitimate equilibrium within the class once substituted into (19) because the fixed-point conditions would be satisfied. These equilibria are indexed by information, since each distinct numerical value of λ reflects how much information is revealed in equilibrium. The notion of “multiplicity” in this scenario is not related to the well-known indeterminacy criteria in rational expectations models, where a continuum of equilibria exists. In

fact, Assumption 1 rules out that type of multiplicity here. Theorem 1 does allow for more than one rational expectations equilibrium in the “exactly one zero” class, and such equilibria are “locally unique” in the sense that small perturbations of the information sets will not lead to an alternative λ -value and therefore will not diverge to an alternative rational expectations equilibrium.

4.2 OUTLINE OF PROOF The proof consists of four steps and can be found in its entirety in Appendix A.2. We briefly discuss each step, relegating tedious algebra to the appendix.

STEP 1: FACTORIZATION We operationalize the key requirement that $Y(\lambda) = 0$ for $\lambda \in (-1, 1)$ by specifying a guess of the form $Y(L) = (L - \lambda)G(L)$, where $G(L)$ has no zeros inside the unit circle. The first step in the proof is to then use the equilibrium guess to derive the canonical factorization for the information set, which can be written as

$$\begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix} = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ (L - \lambda)G(L)\sigma_\varepsilon & 0 \end{bmatrix} \begin{pmatrix} \tilde{\varepsilon}_t \\ \tilde{v}_{it} \end{pmatrix}, \quad (21)$$

where $\varepsilon_t = \sigma_\varepsilon \tilde{\varepsilon}_t$, $v_{it} = \sigma_v \tilde{v}_{it}$, is a convenient normalization so that the variance-covariance matrix of the innovations vector is the identity matrix. The following lemma gives the canonical factorization for $\Gamma(L)$.

Lemma 1. *The canonical factorization $\Gamma^*(z)\Gamma^*(z^{-1})^T$ of the variance-covariance matrix $\Gamma(z)\Gamma(z^{-1})^T$, is given by*

$$\Gamma^*(z) = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}} \begin{bmatrix} A(z)A(\lambda)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v \frac{1-\lambda z}{z-\lambda} (A(z) - A(\lambda)) \\ A(\lambda)\sigma_\varepsilon^2(z - \lambda)G(z) & \sigma_\varepsilon\sigma_v G(z)(1 - \lambda z) \end{bmatrix}. \quad (22)$$

Proof. See Appendix A.2. □

STEP 2: EXPECTATIONS Equipped with the canonical factorization (22), we next derive the three expectational terms: $\mathbb{E}_{it}(x_{it+1})$, $\mathbb{E}_{it}(y_{t+1})$, and $\mathbb{E}_{it}(\theta_{it+1})$ from direct application of the Wiener-Kolmogorov prediction formula. The last two follow directly,

$$\mathbb{E}_{it} \begin{pmatrix} \theta_{it+1} \\ y_{t+1} \end{pmatrix} = [L^{-1}\Gamma^*(L)]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix}.$$

However, the term $\mathbb{E}_{it}(x_{it+1})$, is substantially more involved to derive, due to the fact that the correlation between x_{it+1} and θ_{it} exists not only because they both depend on ε_t , but they also both depend on v_{it} . Formally, the application of the Wiener-Kolmogorov formula leads to

$$\mathbb{E}_{it}(x_{it+1}) = \left[L^{-1}g_{x_i,(\theta_i,y)}(L)(\Gamma^*(L^{-1})^T)^{-1} \right]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix},$$

where $g_{x_i,(\theta_i,y)}(L)$ is the variance-covariance generating function between x_i and the information

set. Given the equilibrium guess, such a function takes the form

$$g_{x_i,(\theta_i,y)}(L) = \begin{bmatrix} X(L)A(L^{-1})\sigma_\varepsilon^2 + V(L)\sigma_v^2 & X(L)(L^{-1} - \lambda)G(L^{-1})\sigma_\varepsilon^2 \end{bmatrix}.$$

A bit of algebra gives

$$L^{-1}g_{x_i,(\theta_i,y)}(L)(\Gamma^*(L^{-1})^T)^{-1} = \begin{bmatrix} L^{-1}(V(L)\sigma_v^2 + X(L)\sigma_\varepsilon^2 A(\lambda)) & \sigma_\varepsilon\sigma_v L^{-1} \frac{1-\lambda L}{L-\lambda} (X(L) - V(L)A(\lambda)) \end{bmatrix}.$$

Acknowledging that the terms have the usual principal part around $L = 0$ and around $L = \lambda$, it follows that

$$\begin{aligned} \mathbb{E}_{it}(x_{it+1}) = & L^{-1} \left[X(L) - X(0) \right] \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] \varepsilon_t \\ & + L^{-1} \left[V(L) - V(0) \right] v_{it} + \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] v_{it}. \end{aligned} \quad (23)$$

STEP 3: FIXED POINT. Next, we need to derive and check the fixed-point conditions. This amounts to algebraic manipulations that serve to get the model in the form such that existence and uniqueness criteria can be invoked, as well as the condition $Y(\lambda) = 0$. Here we report the part of the proof that focuses on making sure that the fixed point in information is established. The proof consists in checking that when the equilibrium coefficients are evaluated using a λ that solves (20), there are no other points at which $Y(L)$ vanishes inside the unit circle. More precisely, it has to be that there is no $\xi \neq \lambda$ that solves

$$\mathcal{Y}(\xi)\tilde{\Phi}(\xi) = (1 - \tau(\lambda))(1 - \lambda^2) \frac{\mathcal{A}(\lambda)}{1 - \lambda\xi}, \quad (24)$$

such that $|\xi| \in (-1, 1)$. If such a ξ existed, the information conveyed by y_t in equilibrium would be inconsistent with the information used to compute the expectations that are part of the fixed point. To see this, suppose that for a given λ that solves (20), a $|\xi| < 1$ that satisfies (24) exists. The $Y(L)$ solution computed using that λ would have, by construction, another zero at ξ . If we denote that solution by $\tilde{Y}(L) = \tilde{G}(L)(L - \lambda)(L - \xi)$, the factorization (22) would only remove the zero associated with λ so that Step 1 above would give

$$\Gamma^*(L) = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}} \begin{bmatrix} A(L)A(\lambda)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v \frac{1-\lambda L}{L-\lambda} (A(L) - A(\lambda)) \\ A(\lambda)\sigma_\varepsilon^2(L - \lambda)(L - \xi)\tilde{G}(L) & \sigma_\varepsilon\sigma_v \tilde{G}(L)(L - \xi)(1 - \lambda L) \end{bmatrix}, \quad (25)$$

Note that the determinant of $\Gamma^*(L)$ in (25) vanishes at $L = |\xi| < 1$, so the factorization will result in expectations that are conditioned on an information set that is inconsistent with the information revealed in equilibrium. In other words, for the specific λ under consideration, the fixed point in information at the heart of Theorem 1 would fail to be verified.

If no equilibria with confounding dynamics with exactly one zero can be found, one can modify the initial guess and consider $N > 1$ roots inside the unit circle, looking then for a condition analogue to (20) to deliver exactly N solutions. We restrict our attention to $N = 1$ for simplicity

and because the full description of the space of REE with confounding dynamics is beyond the scope of this paper, but we hope it is clear that our methods extend to the more general case.

STEP 4: NO INFORMATION FROM THE MODEL The last thing to check to complete the proof is to ensure that there is no information that is transmitted by a clever manipulation of the model conditions – which are part of the information set of the agents – combined with the knowledge of the history of θ_{it} and y_t . For instance, suppose that the market clearing condition (10) is specified so that $\int_0^1 x_{it}\mu(i)di = y_t$, which means that y_t is the aggregate of x_{it} , then this would imply $X(L) = Y(L)$, which would result in $x_{it} - y_t = V(L)v_{it}$. Because rational agents know this, they know that the difference $x_{it} - y_t$ is just a linear combination of the individual innovations v_{it} . It follows that they could, in principle, back out the realizations of v_{it} 's by inverting $V(L)$. More generally, the link between $X(L)$ and $Y(L)$ due to (10) can be used by rational agents to obtain additional information on the underlying innovations. For this not to happen, if one augments the information set of the agents by $x_{it} - y_t$, the information matrix must still be non-invertible at λ . The following Lemma shows that this is indeed the case for the equilibrium of Theorem 1.

Lemma 2. *In the equilibrium with confounding dynamics of Theorem 1, consider the augmented information matrix $\tilde{\Gamma}(L)$, where*

$$\begin{pmatrix} \theta_{it} \\ y_t \\ x_{it} - y_t \end{pmatrix} = \tilde{\Gamma}(L) \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix} = \begin{bmatrix} A(L) & 1 \\ Y(L) & 0 \\ X(L) - Y(L) & V(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix}. \quad (26)$$

The 2-by-2 minors of $\tilde{\Gamma}(L)$ all vanish at λ .

Proof. See Appendix A.2. □

The form of (19) is intuitive when contrasted with the full information counterpart. The standard Hansen-Sargent formula subtracts off the particular linear combination of *future* values of ε_t that minimize the agent's forecast error. As described in Section 2, confounding dynamics implies that a particular linear combination of *past* values of ε_t are never revealed to the agent. In order to make a direct comparison to the full-information case transparent, set $\gamma_y(L) = \gamma_x(L)$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, $\phi_\theta = 0$ and $\psi_\theta(L) = -1$. According to Theorem 1, the solution under confounding dynamics can be written as

$$y_t = \sum_{j=0}^{\infty} \zeta^j \theta_{t+j} - A(\zeta) \sum_{j=1}^{\infty} \zeta^j \varepsilon_{t+j} - (1 - \tau(\lambda))(1 - \lambda^2)A(\lambda) \sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}. \quad (27)$$

The first two components on the right-hand side of (27) give the standard (full-information) Hansen-Sargent formula. The third component—represented by the weighted sum $\sum_{j=0}^{\infty} \lambda^j \varepsilon_{t-j}$ —arises due to confounding dynamics and is similar to the prediction formula of Section 2. Agents do not observe the linear combination of shocks weighted by λ . Conditioning down implies that this linear combination will (optimally) be subtracted from the Hansen-Sargent full-information equilibrium.

The extent to which the unknown past matters depends on the imprecision of the private signal θ_{it} , measured by $1 - \tau(\lambda)$; the imprecision stemming from confounding dynamics, measured by $1 - \lambda^2$; and the fixed point constant $\mathcal{A}(\lambda)$.

Equation (20) provides the condition for the existence of equilibrium (19). It is obtained by evaluating the right-hand side of (19) at λ and setting it equal to zero. By doing so, (20) is ensuring that once the conditioning down due to confounding dynamics is taken into account, the λ responsible for such conditioning down must indeed be a point in which the equilibrium function is non-invertible. Condition (20) takes an intuitive form from an informational point of view. Note first that the LHS, $\mathcal{Y}(\lambda)\tilde{\Phi}(\lambda)$, corresponds to the moving average part of the full information solution evaluated at λ (a complete derivation of the full-information counterpart is presented in the Appendix A.1). Suppose for a moment that the RHS of (20) is set to zero. If a $|\lambda| \in (0, 1)$ satisfying the condition existed, it would mean that the equilibrium with confounding dynamics would take the same form as the full information equilibrium $\mathcal{Y}(L)$. However, equation (27) shows that in presence of confounding dynamics the unknown past must be subtracted from the full information equilibrium, which would make the full information solution $\mathcal{Y}(L)$ inconsistent with confounding dynamics. The implication of this observation is that whenever the RHS of (20) is made small enough, an equilibrium with confounding dynamics may fail to exist. In particular, as the noise-to-signal ratio in private information $\sigma_v/\sigma_\varepsilon$ declines, the signal-to-noise ratio, $\tau(\lambda)$, gets closer to one, and eventually leads to non-existence of an equilibrium with confounding dynamics.

We finally note that the autoregressive factor in (19), $1/(1 - \lambda L)$, injects into the equilibrium dynamics of y_t the waves of over- and under-reaction or the hump-shaped impulse depicted in Figure 1, which are the hallmark of signal extraction under confounding dynamics. In Section 5, in the context of a real business cycle model, we provide a description of how economic incentives can combine with the signal extraction under non-invertibility to deliver the fixed-point condition (20), and a hump-shaped response to shocks exclusively due to confounding dynamics.

5 APPLICATION: BUSINESS CYCLE WITH CONFOUNDING DYNAMICS

In this section we apply our results to a model of business cycle fluctuations driven by productivity shocks. The purpose of this section is to analytically demonstrate the confounding dynamics mechanism within a well established framework. To achieve this goal, we work within a linearized model reminiscent of the islands model of Lucas (1975). We motivate this section with two observations: First, note that this application allows us to demonstrate that the sufficient conditions for confounding dynamics are non-empty. Second, a common criticism of many models that follow the Lucas tradition is that agents cannot see economy-wide prices: if they could, then they could infer fundamentals perfectly and there would not be any confusion in equilibrium. Our setup does not suffer from this criticism.⁹

The economy consists of a continuum of islands indexed by $i \in [0, 1]$. Each island is inhabited by an infinitely-lived representative household, and by a representative firm, also indexed by i .

⁹We are not unique in this respect, see for instance Amador and Weill (2010).

Household i supplies labor services exclusively to firm i in a decentralized competitive labor market or, equivalently, workers cannot move across islands. Households supply labor inelastically to firms, and the labor supply is normalized to 1. Households own capital in the economy, which is rented out to firms in a centralized spot market. Firms use capital and labor to produce output, also supplied in a centralized competitive spot market. Households derive utility from consuming the output good. Output is produced by firm i according to a Cobb-Douglas technology with capital and labor inputs – with income shares α , and $1 - \alpha$ respectively, and total factor of productivity that is firm-specific and denoted by $e^{a_{it}}$, where

$$a_{it} = a_t + v_{it}.$$

The term a_t is common across all the islands, while v_{it} is a productivity component that is specific to island i . In what follows, we consider a log-linearized version of the model with full capital depreciation and constant elasticity of intertemporal substitution, denoted by $\eta > 0$.¹⁰ Household i sets consumption intertemporally according to the Euler equation

$$\mathbb{E}_{it}(c_{it} - c_{it+1} + \eta r_{t+1}) = 0. \quad (28)$$

The intertemporal budget constraint is

$$(1 - \beta\alpha)c_{it} + \alpha\beta k_{it+1} = (1 - \alpha)w_{it} + \alpha r_t - \alpha k_{it}, \quad (29)$$

where k_{it+1} is the capital stock that household i is carrying into period $t + 1$, w_{it} is the wage rate, r_t is the rental rate of capital, and $\beta \in (0, 1)$ is the subjective discount factor. The island-specific wage rate is given by, $w_{it} = \frac{1}{1-\alpha}(a_{it} - \alpha r_t)$. Aggregate capital is defined as $k_{t+1} \equiv \int_0^1 k_{it+1} \mu(i) di$, and market clearing implies an interest rate

$$r_t = a_t - (1 - \alpha)k_t. \quad (30)$$

Using the household's budget constraint at t and at $t + 1$ to get expressions for c_{it} and c_{it+1} , and leading (30) one period forward, one can substitute (28) into the Euler to obtain a second-order difference equation for capital k_{it+1}

$$\alpha\beta\mathbb{E}_{it}(k_{it+2}) + \eta(1 - \alpha\beta)\mathbb{E}_{it}(r_{t+1}) - \mathbb{E}_{it}(a_{it+1}) = \alpha(1 + \beta)k_{it+1} - \alpha k_{it} - a_{it}, \quad (31)$$

which completely characterizes the equilibrium. As remarked in Section 3.3, the model maps into our general setting by specifying $x_{it} = k_{it+1}$, $y_t = r_t$, and $\theta_{it} = a_{it}$.

Finally, we assume that total factor productivity that is common across islands follows the $AR(1)$ process

$$a_t = \rho a_{t-1} + \varepsilon_t, \quad (32)$$

¹⁰The fully specified model and the derivation of the log-linearization are reported in the Online Appendix B.6.

so that $A(L) = \frac{1}{1-\rho L}$, and with $\rho \in [0, 1]$. Note that there are *no* moving average components in this process, and therefore it is always invertible. It cannot be the source of confounding dynamics. They must emerge naturally from interactions within the model.

FULL INFORMATION We first derive the full information ($\Omega_{it} = v_i^t \vee \varepsilon^t$) solution for aggregate capital and the interest rate. The full-information guess for island-specific capital is given by $k_{it+1} = \mathcal{K}(L)\varepsilon_t + \mathcal{V}(L)v_{it}$. From (30), the interest rate is immediately determined by $r_t = \mathcal{R}(L)\varepsilon_t$ and where

$$\mathcal{R}(L) = A(L) - (1 - \alpha)\mathcal{K}(L)L. \quad (33)$$

The characteristic polynomial associated with equation (31) can be determined as

$$\Phi(L) = \alpha\beta - (\eta(1 - \alpha\beta)(1 - \alpha) + (1 + \beta)\alpha)L + \alpha L^2 = \alpha(\zeta - L)(\beta/\zeta - L). \quad (34)$$

Given that α (capital's share of production) and β (subjective discount factor) are both less than one, (34) contains one root inside the unit circle (ζ) and one outside (β/ζ), and their product is always equal to β . Following the steps outlined in Section A.1, the full information equilibrium for capital can be derived as the $AR(2)$ process

$$\mathcal{K}(L) = \frac{\frac{\zeta}{\alpha\beta}(1 + \kappa)}{(1 - \rho L)(1 - \frac{\zeta}{\beta}L)}, \quad (35)$$

and the interest rate takes an $ARMA(2, 1)$ form

$$\mathcal{R}(L) = \frac{1 - \frac{\zeta}{\alpha\beta}(1 + (1 - \alpha)\kappa)L}{(1 - \rho L)(1 - \frac{\zeta}{\beta}L)}. \quad (36)$$

where $\kappa \equiv \frac{\rho(1-\zeta)(\alpha\beta/\zeta-1)}{(1-\rho\zeta)(1-\alpha)}$.

The standard assumptions in the RBC model imply that productivity affects the interest rate contemporaneously, while investment in capital affects the interest rate with a one period lag. The consequence of this timing assumption is that r_t features a moving average component which we know, from our analysis, can have important informational consequences. Suppose that the moving average in r_t was non-invertible. If agents were asked to extract the history of ε_t based solely on data from r_t , they would face the signal extraction problem described in Section 2. In particular, they would not be able to recover the exact history of ε_t . In the full information equilibrium reported above, agents are assumed to directly observe ε_t in every period, and so the equilibrium dynamics are consistent with the information used to compute expectations even if r_t itself is non-invertible. However, what if we modify the information available to the agents by removing the direct observation of the shocks? How would the equilibrium change? Theorem 1 can be readily applied to address these question, to which we now turn.

CONFOUNDING DYNAMICS The first step in applying Theorem 1 is to specify the agents' infor-

mation set. Because households participate in two competitive markets every period – the labor market and the rental market for capital – they observe the island-specific wage rate w_{it} , and the rental rate r_t . The observation of w_{it} and r_t implies that household i can always back out a_{it} at time t through the expression for w_{it} reported above. As a consequence, observing the prices of labor and capital is equivalent to the information set

$$\Omega_{it} = a_i^t \vee r^t. \quad (37)$$

We also assume that households cannot observe the aggregate capital k_t , so to avoid the full revelation of a_t , and thus v_{it} , which would be implied by (30).¹¹ Following Theorem 1, existence of confounding dynamics requires that the process for $r_t = R(L)\varepsilon_t$, has the following property,

$$R(\lambda) = 0, \quad (38)$$

for a $\lambda \in (-1, 1)$. A direct application of Theorem 1 leads to the following corollary.

Corollary 1. *Consider the Real Business Cycle model (30)-(32). Let the information sets be specified as in (37). There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form, $k_{t+1} = K(L)\varepsilon_t$, and $r_t = R(L)\varepsilon_t$, with*

$$K(L) = \mathcal{K}(L) - (1 - \tau(\lambda))\mathcal{C}(\lambda) \frac{(1 - \frac{\zeta}{\beta}\lambda)(1 - \lambda^2)}{(1 - \frac{\zeta}{\beta}L)(1 - \lambda L)}, \quad (39)$$

and $R(L) = A(L) - (1 - \alpha)K(L)L$, if there exists a $\lambda \in (-1, 1)$, that solves

$$\mathcal{R}(\lambda) = -(1 - \tau(\lambda))(1 - \alpha)\mathcal{C}(\lambda)\lambda, \quad (40)$$

where $\mathcal{C}(\lambda) \equiv \frac{(\frac{1-\beta}{1-\rho\lambda})((1-\lambda\beta)\lambda - \tau(\lambda)(1-\lambda^2)\beta) + (1-\tau(\lambda))(1-\lambda^2)\beta((1-\rho\zeta)\kappa/\rho - \zeta) + (1-\lambda\beta)\lambda\kappa/\rho}{\lambda\alpha(\lambda - \beta/\zeta)((1-\lambda\beta)(\lambda - \zeta) - (1-\lambda^2)\tau(\lambda)(\beta - \zeta))}$. $\mathcal{K}(L)$ and $\mathcal{R}(L)$ are as in (35) and (36), $\tau(\lambda) \equiv \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + (1-\rho\lambda)^2\sigma_v^2}$, and $R(L)$ has a zero inside the unit circle equal to λ .

While the functional forms of equations (39)–(40) have the same general structure as Theorem 1 (and same interpretation), the context of the application allows us to gain additional insights into the existence and behavior of an equilibrium with confounding dynamics.¹²

Table 1 reports the endogenous values of λ computed solving (40); “none” indicates that there is no $\lambda \in (-1, 1)$ that solves (40). In Panel 1, the elasticity of substitution, η , is held fixed at

¹¹There are many other information structures that would preserve confounding dynamics in this setting and would be consistent with the general specification of Section 3.1.

¹²Note that, as it is the case for Theorem 1, the corollary looks for an equilibrium functional form with exactly one root (λ) inside the unit circle. In our numerical analysis, it can happen that condition (40) is satisfied by more than one numerical value for λ . For each individual numerical value, we verify that the initial conjectured equilibrium functional form holds, so we confirm that we have identified a rational expectations equilibrium (i.e. we implement the fixed point check described in Step 3 of the sketch of the proof of Theorem 1). We have also verified that the qualitative properties of the equilibrium are the same across numerical values and the overall message of our results do not change.

Table 1: Existence of Equilibrium with Confounding Dynamics

Panel 1: $\eta = 1$	(a)	(b)	(c)
Noise-Signal Ratio, $\sigma_v/\sigma_\varepsilon$	0.1	1	4
Confounding Dynamics, λ	none	none	0.73
Panel 2: $\sigma_v/\sigma_\varepsilon = 2$	(a)	(b)	(c)
Elasticity of Substitution, η	0.5	1	2
Confounding Dynamics, λ	0.44	none	none

Existence of Equilibrium with Confounding Dynamics for numerical values of the noise-to-signal ratio in a_{it} , $\sigma_v/\sigma_\varepsilon$, and the elasticity of intertemporal substitution, η . The rest of the parameters are set at $\beta = 0.985$, $\alpha = .33$, $\rho \approx 1$. The entry “none” indicates that there is no $\lambda \in (-1, 1)$ that solves (40).

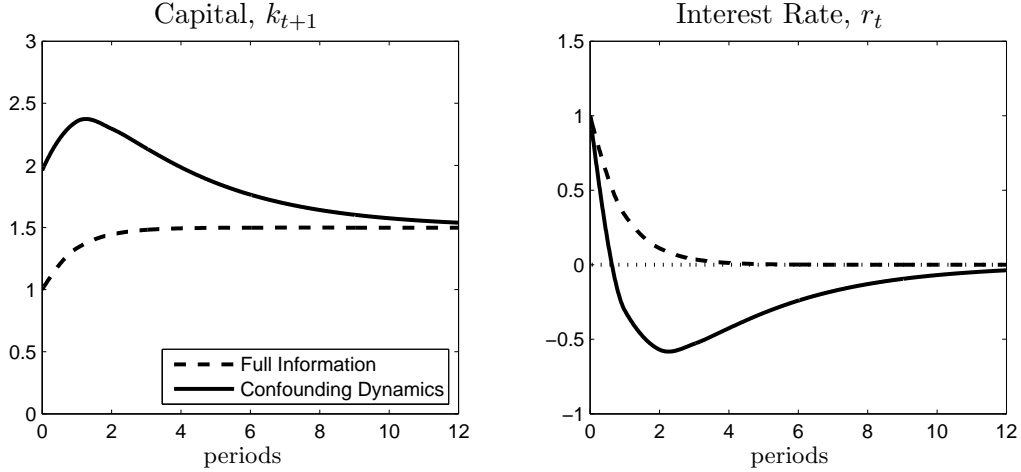
1 – corresponding to log utility – and the private signal precision, $\sigma_v/\sigma_\varepsilon$, is changed from very informative (column (a)) to very uninformative (column (c)). An equilibrium with confounding dynamics exists when the private signal is uninformative: column (c) with $\lambda = 0.73$. Intuitively, if the private signal is very informative, agents will rely strongly on their private information in forming their beliefs about aggregate productivity, which, in turn, will make the interest rate more informative. In Panel 2, $\sigma_v/\sigma_\varepsilon$ is held fixed at 2, and the elasticity of substitution is changed from a low level (0.5 in column (a)), to a high level (2 in column (c)). In this case, the equilibrium with confounding dynamics only exists when the elasticity of substitution is sufficiently low: column (a), with $\lambda = 0.44$. From the full information equilibrium (35) we see that a lower elasticity of substitution implies a more sluggish response of capital as agents are less willing to substitute consumption for investment.¹³ In the presence of incomplete information, a similar sluggish adjustment prevents capital, and thus the interest rate, to correctly reflect the underlying changes in fundamentals. As the elasticity of substitution is increased, the more reactive response of capital results in the interest rate dynamics fully revealing the fundamentals.

We conclude the analysis by using the case in column (c), in Panel 1 of Table 1 to study the qualitative effects of confounding dynamics on capital and the interest rate. Figure 2 shows the response of capital, k_{t+1} , and the interest rate, r_t , to a persistent unitary positive shock to aggregate productivity a_t under full information (dashed lines) and confounding dynamics (plain lines). Under full information, capital increases at impact and steadily climbs towards a new persistent level (recall that $\rho \approx 1$ in this example). The interest rate increases at impact because capital is fixed at first while productivity is higher. Subsequently, the interest rate steadily declines because of the increased capital accumulation which reduces the marginal product of capital.

In the equilibrium with confounding dynamics the impulse responses are markedly different. As shown by the solid lines in Figure 2, the response of capital is amplified for every t and displays

¹³To see this, note that (34) implies $\frac{\partial \zeta}{\partial \eta} > 0$, which in turn implies $\frac{\partial \mathcal{K}(0)}{\partial \eta} > 0$ for $\rho \approx 1$.

Figure 2: Impulse Response of Capital and Interest Rate



Impulse response of Capital, k_{t+1} , and Interest Rate, r_t , under Full Information (dash-line) and Confounding Dynamics (plain-line). The parameter values are $\eta = 1$, $\beta = 0.985$, $\alpha = 0.33$, $\rho \approx 1$, and $\sigma_v/\sigma_e = 4$. For the Confounding Dynamics equilibrium, $\lambda = 0.73$.

a hump-shaped pattern, with the peak reached in period 1 and a persistent slow decline towards the full information long-run level. The interest rate at impact is equal to the full-information case because capital is fixed, but it drops in negative territory in the subsequent periods because of the larger response of capital.

The intuition as of why capital displays an amplified response under confounding dynamics can be found in how capital behaves to ensure that the interest rate is non-invertible at λ . When $\eta = 1$, and $\rho \approx 1$, one can show that ¹⁴

$$\mathcal{R}(L) \approx \frac{1}{(1 - \alpha L)}, \quad (41)$$

which means that the full information interest rate decays gradually towards zero after impact as capital gradually climbs towards a very persistent higher level. In the equilibrium with confounding dynamics, a moving average component with root λ appears in the process for the interest rate, which requires the interest rate to overshoot into negative territory after impact, similarly to Panel B in Figure 1. For that to be the case, the interest rate must decline more than the full-information case one period after impact, which, given the timing of the model, can happen only with a higher response of capital at impact. Using (41) together with (40) it is possible to show that the difference between the dynamic response of capital across the two equilibria is

$$K(L) - \mathcal{K}(L) \approx \frac{(1 - \lambda^2)}{(1 - \alpha)\lambda(1 - \alpha L)(1 - \lambda L)}. \quad (42)$$

¹⁴To see this note that when $\eta = 1$, $\zeta/\beta = \alpha$, and $\kappa = 0$, so the expression immediately follows from (36) when one recognizes that the ratio $(1 - L)(1 - \rho L)$ cancels for $\rho \approx 1$. Also note that this relationship only holds under full information.

The hump shape with peak at period 1 emerges because, $\alpha + \lambda > 1$, and, $\alpha^2 + \lambda^2 + \alpha\lambda < \alpha + \lambda$, in our numerical example. Intuitively, the persistence in the interest rate dynamics, measured by α , combines with the persistence due to signal extraction from the interest rate, measured by λ , and they initially reinforce each other before eventually declining.

The role of the informativeness of private signal, a_{it} , measured by $\sigma_v/\sigma_\varepsilon$, is also crucial in sustaining the amplified response and thus non-invertibility. Based just on their private signal, optimal signal extraction would instruct agents to be conservative in estimating the innovation to aggregate productivity, which would result in lower investment at the individual agent’s level compared to full information and thus aggregate under-reaction of investment. However, the behavior of the interest rate under confounding dynamics changes the average *predicted* innovation in a_t . If agents observe a large drop in the interest rate after impact, their signal extraction effort leads them to rationally infer that the aggregate productivity shock is larger than what their private signal alone would suggest. In this sense, the interest rate dynamics, when used to extract information about the innovation in productivity, acts as a perceived positive *aggregate* innovation in productivity. If the private signal is sufficiently uninformative, the perceived innovation remains consistent with rational expectations, and an equilibrium with confounding dynamics is established.

Our application starkly showcases the central insight coming from Theorem 1: allowing for the endogeneity of signals in a dynamic context opens the door to a set of equilibria that are usually overlooked when information is exogenously provided to the agents. Figure 2 shows that equilibria with confounding dynamics can display a qualitative behavior of key aggregate variables that is interesting and promising for quantitative applications. The shape and size of the response is determined by the assumption that we look at equilibria with only one non-invertible root λ . However, richer non-invertible conditions – such as ones with multiple roots, conjugate pairs, etc. – would result in richer dynamics that would ensure a better fit of data (we explore a simple prediction example with multiple roots in Appendix B.4). Finally, in order to keep things analytically tractable and transparent, we have assumed away additional sources of frictions, thereby limiting the potential of the model to provide quantitatively significant results. However, we envision a richer environment with several types of frictions, such as financial frictions – which are likely to introduce stronger sensitivity of allocations to the interest rate, or exogenous noisy signals, but where confounding dynamics remain a major determinant of equilibrium behavior.

6 CONCLUDING COMMENTS

As we have shown, confounding dynamics injects persistence into impulse response functions. These interesting dynamics are generated from a simple and optimal learning mechanism that can be easily applied to any dynamic setting. Future work will seek to better understand the empirical properties of confounding dynamics by incorporating them into real and nominal business cycle models designed to be taken to data. Theoretical results of Section 5 and preliminary empirical results show much promise. Future work will also seek to show an equivalence between the analytic function approach advocated here and the more familiar time-domain approach. Contrasting these

approaches in a side-by-side fashion will help to highlight the benefits of the analytic function approach while demystifying certain aspects of it.

REFERENCES

- ACHARYA, S., J. BENHABIB, AND Z. HUO (2017): “The Anatomy of Sentiment-Driven Fluctuations,” Working Paper 23136, National Bureau of Economic Research.
- ADAM, K. (2007): “Optimal Monetary Policy with Imperfect Common Knowledge,” *Journal of Monetary Economics*, 54(2), 267–301.
- ALLEN, F., S. MORRIS, AND H. SHIN (2006): “Beauty contests and iterated expectations in asset markets,” *Review of Financial Studies*, 19(3), 719–752.
- AMADOR, M., AND P.-O. WEILL (2010): “Learning from Prices: Public Communication and Welfare,” *Journal of Political Economy*, 118(5), 866–907.
- ANGELETOS, G.-M., AND J. LA’O (2009): “Noisy Business Cycles,” *NBER Macroeconomics Annual*, 24.
- ANGELETOS, G.-M., AND J. LA’O (2013): “Sentiments,” *Econometrica*, 81(2), 739–780.
- ANGELETOS, G.-M., AND C. LIAN (2016): “Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination,” *Handbook of Macroeconomics*, 2, 1065–1240.
- ANGELETOS, G.-M., AND A. PAVAN (2007): “Efficient Use of Information and Social Value of Information,” *Econometrica*, 75(4).
- BACCHETTA, P., AND E. VAN WINCOOP (2006): “Can Information Heterogeneity Explain the Exchange Rate Puzzle?,” *American Economic Review*, 96(3), 552–576.
- BENHABIB, J., P. WANG, AND Y. WEN (2015): “Sentiments and Aggregate Demand Fluctuations,” *Econometrica*, 83(2), 549–585.
- BERNHARDT, D., P. SEILER, AND B. TAUB (2010): “Speculative Dynamics,” *Economic Theory*, 44, 1–52.
- BLANCHARD, O. J., AND C. M. KAHN (1980): “The Solution of Linear Difference Models under Rational Expectations,” *Econometrica*, 48(5), 1305–1312.
- BROCKWELL, P. J., AND R. A. DAVIS (1987): *Time Series: Theory and Methods*. Springer Series in Statistics, Springer-Verlag New York.
- BROWN, J., AND R. CHURCHILL (2013): *Complex Variables and Applications*. McGraw-Hill Education, New York, USA, ninth edn.

- CAGAN, P. D. (1956): “The Monetary Dynamics of Hyperinflation,” in *Studies in the Quantity Theory of Money*, ed. by M. Friedman. University of Chicago Press.
- CANOVA, F., AND M. H. SAHNEH (2017): “Are Small-Scale SVARs Useful for Business Cycle Analysis? Revisiting Nonfundamentalness,” *Journal of the European Economic Association*, 16(4), 1069–1093.
- COGLEY, T., AND J. M. NASON (1995): “Output Dynamics in Real-Business-Cycle Models,” *American Economic Review*, 85(3), 492–511.
- COIBION, O., AND Y. GORODNICHENKO (2012): “What Can Survey Forecasts Tell Us about Information Rigidities?,” *Journal of Political Economy*, 120(1), 116–159.
- FERNANDEZ-VILLAYERDE, J., J. F. RUBIO, T. SARGENT, AND M. WATSON (2007): “A,B,C (and D)’s for Understanding VARs,” *American Economic Review*, 97, 1021–1026.
- FUTIA, C. A. (1981): “Rational Expectations in Stationary Linear Models,” *Econometrica*, 49(1), 171–192.
- GALÌ, J. (2008): *Monetary Policy, Inflation and the Business Cycle: An Introduction to the New Keynesian Framework*. Princeton University Press, Princeton and Oxford.
- GRAHAM, L., AND S. WRIGHT (2010): “Information, Heterogeneity and Market Incompleteness,” *Journal of Monetary Economics*, 57(2), 164–174.
- GREGOIR, S., AND P. WEILL (2007): “Restricted Perception Equilibria and Rational Expectation Equilibrium,” *Journal of Economic Dynamics and Control*, 31(1), 81–109.
- GROSSMAN, S. J., AND J. E. STIGLITZ (1980): “On the Impossibility of Informationally Efficient Markets,” *American Economic Review*, 70, 393–408.
- HANSEN, L. P., AND T. J. SARGENT (1980): “Formulating and Estimating Dynamic Linear Rational Expectations Models,” *Journal of Economic Dynamics and Control*, 2, 7–46.
- HASSAN, T. A., AND T. M. MERTENS (2011): “The Social Cost of Near-Rational Investment,” NBER Working Paper 17027.
- HELLWIG, C. (2006): “Monetary Business Cycle Models: Imperfect Information,” *New Palgrave Dictionary of Economics*.
- HELLWIG, C., AND V. VENKATESWARAN (2009): “Setting the Right Prices for the Wrong Reasons,” *Journal of Monetary Economics*, 56, S57–S77.
- HOFFMAN, K. (1962): *Banach Spaces of Analytic Functions*. Prentice-Hall, Englewood Cliffs, New Jersey.

- HUERTGEN, P., M. HOFFMANN, G. RONDINA, AND T. B. WALKER (2016): “Rational Expectations with Dispersed Information: A Comparison of Solution Techniques,” CAEPR Working Paper.
- HUO, Z., AND N. TAKAYAMA (2016): “Rational Expectations Models with Higher Order Beliefs,” Working Paper.
- KASA, K. (2000): “Forecasting the Forecasts of Others in the Frequency Domain,” *Review of Economic Dynamics*, 3, 726–756.
- KASA, K., T. B. WALKER, AND C. H. WHITEMAN (2014): “Heterogenous Beliefs and Tests of Present Value Models,” *Review of Economic Studies*, 81(3), 1137–1163.
- KEYNES, J. M. (1936): *The General Theory of Employment, Interest and Money*. Macmillan, London, U.K.
- KING, R. (1982): “Monetary Policy and the Information Content of Prices,” *Journal of Political Economy*, 90(2), 247–279.
- 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.
- KLENOW, P. J., AND J. L. WILLIS (2007): “Sticky information and sticky prices,” *Journal of Monetary Economics*, 54, 79 – 99.
- LORENZONI, G. (2009): “A Theory of Demand Shocks,” *American Economic Review*, 99(5), 2050–2084.
- LUCAS, JR., R. E. (1972): “Expectations and the Neutrality of Money,” *Journal of Economic Theory*, 4, 103–124.
- (1975): “An Equilibrium Model of the Business Cycle,” *Journal of Political Economy*, 83, 1113–1144.
- MAKAROV, I., AND O. RYTCHKOV (2012): “Forecasting the Forecasts of Others: Implications for Asset Pricing,” *Journal of Economic Theory*, 147, 941–966.
- MANKIW, N., AND R. REIS (2002): “Sticky Information Versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 117(4), 1295–1328.
- MELOSI, L. (2016): “Signaling Effects of Monetary Policy,” forthcoming, *Review of Economic Studies*.
- NIMARK, K. (2008): “Dynamic Pricing and Imperfect Common Knowledge,” *Journal of Monetary Economics*, (55), 365–382.

- NIMARK, K. (2010): “Dynamic higher order expectations,” Working paper, Universitat Pompeu Fabra.
- PEARLMAN, J. G., AND T. J. SARGENT (2005): “Knowing the Forecasts of Others,” *Review of Economic Dynamics*, 8(2), 480–497.
- PIGOU, A. C. (1929): *Industrial Fluctuations*. Macmillan, London, second edn.
- RADNER, R. (1979): “Rational Expectations Equilibrium: Generic Existence and the Information Revealed by Prices,” *Econometrica*, 47(3), 655–678.
- RONDINA, G. (2009): “Incomplete Information and Informative Pricing,” Working Paper. UCSD.
- ROZANOV, Y. A. (1967): *Stationary Random Processes*. Holden-Day, San Francisco.
- SARGENT, T. (1987): *Macroeconomic Theory*. Emerald Group Publishing Limited, Bingley, UK, second edn.
- SIMS, C. A. (2002): “Solving Linear Rational Expectations Models,” *Computational Economics*, 20, 1–20.
- SINGLETON, K. J. (1987): “Asset Prices in a Time Series Model with Disparately Informed, Competitive Traders,” in *New Approaches to Monetary Economics*, ed. by W. Barnett, and K. Singleton. Cambridge University Press, Cambridge.
- TAUB, B. (1989): “Aggregate Fluctuations as an Information Transmission Mechanism,” *Journal of Economic Dynamics and Control*, 13(1), 113–150.
- TOWNSEND, R. M. (1983a): “Equilibrium Theory with Learning and Disparate Expectations: Some Issues and Methods,” in *Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined*, ed. by E. S. Phelps, and R. Frydman, pp. 169–202. Cambridge University Press.
- (1983b): “Forecasting the Forecasts of Others,” *Journal of Political Economy*, 91(4), 546–588.
- UHLIG, H. (1999): “Analyzing Non Linear Dynamics Stochastic Models,” in *Computational Methods for the Study of Dynamic Economies*, ed. by A. S. R. Marimon, pp. 30–61. Oxford University Press, Oxford.
- WALKER, T. B. (2007): “How Equilibrium Prices Reveal Information in Time Series Models with Disparately Informed, Competitive Traders,” *Journal of Economic Theory*, 137(1), 512–537.
- WANG, J. (1993): “A Model of Asset Prices under Asymmetric Information,” *Review of Economic Studies*, 60(2), 249–282.

- WHITEMAN, C. (1983): *Linear Rational Expectations Models: A User's Guide*. University of Minnesota Press, Minneapolis.
- WHITTLE, P. (1983): *Prediction and Regulation by Linear Least-Square Methods*. University of Minnesota Press, Minneapolis, second edn.
- WOODFORD, M. (2003a): "Imperfect Common Knowledge and the Effects of Monetary Policy," in *Knowledge, Information, and Expectations in Modern Macroeconomics*, ed. by P. Aghion, R. Frydman, J. Stiglitz, and M. Woodford. Princeton University Press, Princeton, N.J.
- (2003b): *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ.

A PROOFS

A.1 FULL INFORMATION SOLUTION The proof of Theorem 1 makes use of the full information solution of (10)-(7). We report the derivation of the full information solution here for completeness. We define as Full Information the case when every agent is endowed with perfect knowledge of the aggregate and her own idiosyncratic innovations history up to time t . Denoting the full information set by $\tilde{\Omega}_{it}$, the set is formally specified as

$$\tilde{\Omega}_{it} = v_i^t \vee \varepsilon^t. \quad (\text{A.1})$$

Here, and in the following analysis, we assume that agents know that the equilibrium relationship is given by (7)-(10). We begin by guessing that the solution takes the form, $x_{it} = \mathcal{X}(L)\varepsilon_t + \mathcal{V}(L)v_{it}$, and $y_t = \mathcal{Y}(L)\varepsilon_t$, where $\mathcal{X}(L)$, $\mathcal{V}(L)$ and $\mathcal{Y}(L)$ are square-summable lag polynomial in non-negative powers of L . Under full information, direct application of the Wiener-Kolmogorov formula (see Appendix C) provides expressions for the relevant expectational terms,

$$\mathbb{E}_{it}(x_{it+1}) = [\mathcal{X}(L) - \mathcal{X}(0)]L^{-1}\varepsilon_t + [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it}, \quad (\text{A.2})$$

$$\mathbb{E}_{it}(y_{t+1}) = [\mathcal{Y}(L) - \mathcal{Y}(0)]L^{-1}\varepsilon_t, \quad (\text{A.3})$$

$$\mathbb{E}_{it}(\theta_{t+1}) = [A(L) - A(0)]L^{-1}\varepsilon_t. \quad (\text{A.4})$$

The fixed point condition under full information can be found by substituting (A.2)-(A.4) into (7), so that

$$\begin{aligned} & \phi_x [\mathcal{X}(L) - \mathcal{X}(0)]L^{-1}\varepsilon_t + \phi_x [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it} + \phi_y [\mathcal{Y}(L) - \mathcal{Y}(0)]L^{-1}\varepsilon_t + \phi_\theta [A(L) - A(0)]L^{-1}\varepsilon_t \\ & = \psi_x(L)\mathcal{X}(L)\varepsilon_t + \psi_x(L)\mathcal{V}(L)v_{it} + \psi_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}. \end{aligned} \quad (\text{A.5})$$

This equation defines a fixed point condition for $\mathcal{V}(L)$ with all the terms that multiply v_{it} . Collecting terms that multiply v_{it} , multiplying both sides by L and rearranging we get

$$\mathcal{V}(L)(\phi_x - \psi_x(L)L) = \phi_x\mathcal{V}(0) + \psi_\theta(L)L. \quad (\text{A.6})$$

Note that $\phi_x(L) \equiv \phi_x - \psi_x(L)L$, which, under Assumption 1, has exactly one zero inside the unit circle, which we term ζ_x . We thus pick $\mathcal{V}(0)$ to remove such zero by setting

$$\phi_x\mathcal{V}(0) + \psi_\theta(\zeta_x)\zeta_x = 0. \quad (\text{A.7})$$

Solving for $\mathcal{V}(0)$, substituting back into (A.6) one finally obtains

$$\mathcal{V}(L) = \frac{\psi_\theta(L)L - \psi_\theta(\zeta_x)\zeta_x}{\phi_x(L)}. \quad (\text{A.8})$$

We now focus on the fixed point for $\mathcal{Y}(L)$ and $\mathcal{X}(L)$. As remarked in the text, the fixed point condition does not feature any components of $\mathcal{V}(L)$, so that one does not need to solve for the latter to obtain the former. To proceed with the solution there are two possibilities: solve for $\mathcal{Y}(L)$ and then recover $\mathcal{X}(L)$, or viceversa. In general, both routes are possible, but there are situations in which one direction is substantially easier than the other. This depends on whether $\gamma_x(0) \neq 0$ or $\gamma_y(0) \neq 0$. We report here both cases. We first consider the case that works whenever $\gamma_x(0) \neq 0$. We begin by manipulating condition (10) to get the following relationship between $\mathcal{X}(L)$ and $\mathcal{Y}(L)$,

$$\mathcal{X}(L) = \tilde{\gamma}_y(L)\mathcal{Y}(L) + \tilde{\gamma}_\theta(L)A(L), \quad (\text{A.9})$$

where $\tilde{\gamma}_y(L) = -\frac{\gamma_y(L)}{\gamma_x(L)}$, and $\tilde{\gamma}_\theta(L) = -\frac{\gamma_\theta(L)}{\gamma_x(L)}$. Using (A.9) to substitute for terms featuring $\mathcal{X}(L)$ in (A.5) one obtains

$$\begin{aligned} & \phi_x [\tilde{\gamma}_y(L)\mathcal{Y}(L) - \tilde{\gamma}_y(0)\mathcal{Y}(0)]L^{-1}\varepsilon_t + \phi_x [\tilde{\gamma}_\theta(L)A(L) - \tilde{\gamma}_\theta(0)A(0)]L^{-1}\varepsilon_t + \phi_x [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it} \\ & + \phi_y [\mathcal{Y}(L) - \mathcal{Y}(0)]L^{-1}\varepsilon_t + \phi_\theta [A(L) - A(0)]L^{-1}\varepsilon_t = \psi_x(L)\tilde{\gamma}_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_x(L)\tilde{\gamma}_\theta(L)A(L)\varepsilon_t \\ & + \psi_x(L)\mathcal{V}(L)v_{it} + \psi_y(L)\mathcal{Y}(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}. \end{aligned} \quad (\text{A.10})$$

Taking all the terms that multiply ε_t in (A.10), multiplying by L both sides and rearranging, one gets

$$\mathcal{Y}(L)\Phi(L) = \mathcal{Y}(0)(\phi_x\tilde{\gamma}_y(0) + \phi_y) - \xi_y(L), \quad (\text{A.11})$$

where

$$\xi_y(L) \equiv (\phi_x - \psi_x(L)L)\tilde{\gamma}_\theta(L)A(L) + (\phi_\theta - \psi_\theta(L)L)A(L) - (\phi_x\tilde{\gamma}_\theta(0) + \phi_\theta)A(0). \quad (\text{A.12})$$

Under Assumption 1, $\Phi(L)$ has exactly one zero inside the unit circle, denoted by ζ , which means that we can choose $\mathcal{Y}(0)$ to remove such zero. We thus set

$$\mathcal{Y}(0)(\phi_x\tilde{\gamma}_y(0) + \phi_y) - \xi_y(\zeta) = 0. \quad (\text{A.13})$$

Solving for $\mathcal{Y}(0)$, substituting into (A.11) and rearranging, one finally gets

$$\mathcal{Y}(L) = \frac{\xi_y(\zeta) - \xi_y(L)}{\Phi(L)}. \quad (\text{A.14})$$

The expression for $\mathcal{X}(L)$ can then be recovered using (A.9). Next we consider the case that works whenever $\gamma_y(0) \neq 0$. We begin by manipulating condition (10) to get the following relationship between $\mathcal{X}(L)$ and $\mathcal{Y}(L)$,

$$\mathcal{Y}(L) = \hat{\gamma}_x(L)\mathcal{X}(L) + \hat{\gamma}_\theta(L)A(L), \quad (\text{A.15})$$

where $\hat{\gamma}_x(L) = -\frac{\gamma_x(L)}{\gamma_y(L)}$, and $\hat{\gamma}_\theta(L) = -\frac{\gamma_\theta(L)}{\gamma_y(L)}$. Using (A.15) to substitute for terms featuring $\mathcal{Y}(L)$ in (A.5) one obtains

$$\begin{aligned} & \phi_x [\mathcal{X}(L) - \mathcal{X}(0)]L^{-1}\varepsilon_t + \phi_x [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it} + \phi_y [\hat{\gamma}_x(L)\mathcal{X}(L) - \hat{\gamma}_x(L)\mathcal{X}(0)]L^{-1}\varepsilon_t \\ & + \phi_y [\hat{\gamma}_\theta(L)A(L) - \hat{\gamma}_\theta(L)A(0)]L^{-1}\varepsilon_t + \phi_\theta [A(L) - A(0)]L^{-1}\varepsilon_t = \psi_x(L)\mathcal{X}(L)\varepsilon_t + \psi_x(L)\mathcal{V}(L)v_{it} + \psi_y(L)\hat{\gamma}_x(L)\mathcal{X}(L)\varepsilon_t \\ & + \psi_y(L)\hat{\gamma}_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}. \end{aligned} \quad (\text{A.16})$$

Taking all the terms that multiply ε_t in (A.16), multiplying by L both sides and rearranging, one gets

$$\mathcal{X}(L)\Phi_x(L) = \mathcal{X}(0)(\phi_x + \phi_y\hat{\gamma}_y(0)) - \xi_x(L), \quad (\text{A.17})$$

where

$$\Phi_x(L) = \phi_x + \phi_y\hat{\gamma}_x(L) - \psi_x(L)L - \psi_y(L)\hat{\gamma}_x(L)L, \quad (\text{A.18})$$

and

$$\xi_x(L) \equiv (\phi_y - \psi_y(L)L)\hat{\gamma}_\theta(L)A(L) + (\phi_\theta - \psi_\theta(L)L)A(L) - (\phi_y\hat{\gamma}_\theta(0) + \phi_\theta)A(0). \quad (\text{A.19})$$

Analogously to Assumption 1, let us assume that $\Phi_x(L)$ has exactly one zero inside the unit circle, denoted by $\hat{\zeta}$, which means that we can choose $\mathcal{X}(0)$ to remove such zero. We thus set

$$\mathcal{X}(0)(\phi_x + \phi_y\hat{\gamma}_y(0)) - \xi_x(\hat{\zeta}) = 0. \quad (\text{A.20})$$

Solving for $\mathcal{X}(0)$, substituting into (A.17) and rearranging, one finally gets

$$\mathcal{X}(L) = \frac{\xi_x(\hat{\zeta}) - \xi_x(L)}{\Phi_x(L)}. \quad (\text{A.21})$$

The expression for $\mathcal{Y}(L)$ can then be recovered using (A.15).

A.2 PROOF OF THEOREM 1

STEP 1: FACTORIZATION We operationalize the key requirement that $Y(\lambda) = 0$ for $\lambda \in (-1, 1)$ by specifying a guess of the form $Y(L) = (L - \lambda)G(L)$, where $G(L)$ has no zeros inside the unit circle. The first step in the proof is to then use the equilibrium guess to derive the canonical factorization for the information set, so that the Wiener-Kolmogorov formula (see Appendix C) be applied. The information set can be written as

$$\begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix} = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ (L - \lambda)G(L)\sigma_\varepsilon & 0 \end{bmatrix} \begin{pmatrix} \tilde{\varepsilon}_t \\ \tilde{v}_{it} \end{pmatrix}. \quad (\text{A.22})$$

where $\varepsilon_t = \sigma_\varepsilon \tilde{\varepsilon}_t$, $v_{it} = \sigma_v \tilde{v}_{it}$, is a convenient normalization so that the variance-covariance matrix of the innovations vector is the identity matrix. It follows that

$$\Gamma(L) = \begin{bmatrix} A(L)\sigma_\varepsilon & \sigma_v \\ (L - \lambda)G(L)\sigma_\varepsilon & 0 \end{bmatrix}. \quad (\text{A.23})$$

The following Lemma shows the canonical factorization for $\Gamma(L)$.

Lemma A1. *The canonical factorization $\Gamma^*(z)\Gamma^*(z^{-1})^T$ of the variance-covariance matrix $\Gamma(z)\Gamma(z^{-1})^T$, where $\Gamma(z)$ is defined in (A.23), is given by*

$$\Gamma^*(z) = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}} \begin{bmatrix} A(z)A(\lambda)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v \frac{1-\lambda z}{z-\lambda} (A(z) - A(\lambda)) \\ A(\lambda)\sigma_\varepsilon^2(z - \lambda)G(z) & \sigma_\varepsilon\sigma_v G(z)(1 - \lambda z) \end{bmatrix}. \quad (\text{A.24})$$

Proof. Using Rozanov (1967) procedure, $\Gamma^*(z)$ is computed as

$$\Gamma^*(z) = \Gamma(z)W_\lambda B_\lambda(z). \quad (\text{A.25})$$

where

$$W_\lambda = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}} \begin{bmatrix} A(\lambda)\sigma_\varepsilon & -\sigma_v \\ \sigma_v & A(\lambda)\sigma_\varepsilon \end{bmatrix}, \quad \text{and} \quad B_\lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1-\lambda z}{z-\lambda} \end{bmatrix}. \quad (\text{A.26})$$

The form of W_λ is obtained by application of Lemma C1 in Appendix C. Solving out the matrix multiplication after some algebra one obtains (A.24). \square

STEP 2: EXPECTATIONS Equipped with the canonical factorization (A.24), we next derive the three expectational terms: $\mathbb{E}_{it}(x_{it+1})$, $\mathbb{E}_{it}(y_{t+1})$, and $\mathbb{E}_{it}(\theta_{it+1})$ (recall that $\mathbb{E}_{it}(\theta_{it+1}) = \mathbb{E}_{it}(\theta_{t+1})$). The second and third in the list are given by

$$\mathbb{E}_{it} \begin{pmatrix} \theta_{it+1} \\ y_{t+1} \end{pmatrix} = [L^{-1}\Gamma^*(L)]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix}. \quad (\text{A.27})$$

Recalling that $[L^{-1}\Gamma^*(L)]_+ = [\Gamma^*(L) - \Gamma^*(0)]L^{-1}$, and defining $\tau(\lambda) = \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}$ one gets

$$\mathbb{E}_{it}(\theta_{t+1}) = [A(L) - A(0)]L^{-1}\varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} [A(\lambda) - A(0)]\varepsilon_t - \tau(\lambda) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[1 - \frac{A(0)}{A(\lambda)}\right] v_{it}, \quad (\text{A.28})$$

$$\mathbb{E}_{it}(y_{t+1}) = [(L - \lambda)G(L) + \lambda G(0)]L^{-1}\varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{(1-\lambda L)} G(0)\varepsilon_t + \tau(\lambda) \frac{1-\lambda^2}{(1-\lambda L)} \frac{G(0)}{A(\lambda)} v_{it}. \quad (\text{A.29})$$

The term $\mathbb{E}_{it}(x_{it+1})$, is substantially more involved to derive, due to the fact that the correlation between x_{it+1} and θ_{it} exists not only because they both depend on ε_t , but they also both depend on v_{it} . Formally, the application of the Wiener-Kolmogorov formula leads to

$$\mathbb{E}_{it}(x_{it+1}) = \left[L^{-1}g_{x_i,(\theta_i,y)}(L)(\Gamma^*(L^{-1})^T)^{-1} \right]_+ \Gamma^*(L)^{-1} \begin{pmatrix} \theta_{it} \\ y_t \end{pmatrix}, \quad (\text{A.30})$$

where $g_{x_i,(\theta_i,y)}(L)$ is the variance-covariance generating function between x_i and the information set. Given the

equilibrium guess, such function takes the form

$$g_{x_i,(\theta_i,y)}(L) = \begin{bmatrix} X(L)A(L^{-1})\sigma_\varepsilon^2 + V(L)\sigma_v^2 & X(L)(L^{-1} - \lambda)G(L^{-1})\sigma_\varepsilon^2 \end{bmatrix}. \quad (\text{A.31})$$

It follows that

$$L^{-1}g_{x_i,(\theta_i,y)}(L)(\Gamma^*(L^{-1})^T)^{-1} = \begin{bmatrix} L^{-1}(V(L)\sigma_v^2 + X(L)\sigma_\varepsilon^2 A(\lambda)) & \sigma_\varepsilon \sigma_v L^{-1} \frac{1-\lambda L}{L-\lambda} (X(L) - V(L)A(\lambda)) \end{bmatrix}. \quad (\text{A.32})$$

The application of the annihilator operator requires to take the annihiland minus the principal part of its Laurent series expansion. All the terms have the usual principal part around $L = 0$. However, the term containing $\frac{1-\lambda L}{L-\lambda}$ also has a principal part around $L = \lambda$, it follows that

$$\begin{aligned} & \left[\left(\frac{1-\lambda L}{L-\lambda} \right) \frac{1}{L} (X(L) - V(L)A(\lambda)) \right]_+ \\ & = L^{-1} \left[\left(\frac{1-\lambda L}{L-\lambda} \right) (X(L) - V(L)A(\lambda)) + \frac{1}{\lambda} (X(0) - V(0)A(\lambda)) \right] - \frac{1-\lambda^2}{L-\lambda} \frac{1}{\lambda} (X(\lambda) - V(\lambda)A(\lambda)). \end{aligned} \quad (\text{A.33})$$

Finally one gets

$$\begin{aligned} \mathbb{E}_{it}(x_{it+1}) &= L^{-1} \left[X(L) - X(0) \right] \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] \varepsilon_t \\ & \quad + L^{-1} \left[V(L) - V(0) \right] v_{it} + \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] v_{it}. \end{aligned} \quad (\text{A.34})$$

STEP 3: FIXED POINT We begin by manipulating condition (10) to get the following relationship between $X(L)$ and $Y(L)$,

$$X(L) = \tilde{\gamma}_y(L)Y(L) + \tilde{\gamma}_\theta(L)A(L), \quad (\text{A.35})$$

where $\tilde{\gamma}_y(L) = -\frac{\gamma_y(L)}{\gamma_x(L)}$, and $\tilde{\gamma}_\theta(L) = -\frac{\gamma_\theta(L)}{\gamma_x(L)}$. Next we substitute the equilibrium guess and expressions (A.28), (A.29), and (A.34) into model (7), which leads to the expression

$$\begin{aligned} & \phi_x \left[L^{-1} \left[X(L) - X(0) \right] \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] \varepsilon_t \right. \\ & \quad \left. + L^{-1} \left[V(L) - V(0) \right] v_{it} + \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] v_{it} \right] \\ & \quad + \phi_y \left[[(L - \lambda)G(L) + \lambda G(0)] L^{-1} \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{(1-\lambda L)} G(0) \varepsilon_t + \tau(\lambda) \frac{1-\lambda^2}{(1-\lambda L)} \frac{G(0)}{A(\lambda)} v_{it} \right] \\ & \quad + \phi_\theta \left[[A(L) - A(0)] L^{-1} \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} [A(\lambda) - A(0)] \varepsilon_t - \tau(\lambda) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[1 - \frac{A(0)}{A(\lambda)} \right] v_{it} \right] \\ & = \psi_x(L)(X(L)\varepsilon_t + V(L)v_{it}) + \psi_y(L)(L - \lambda)G(L)\varepsilon_t + \psi_\theta(L)A(L)\varepsilon_t + \psi_\theta(L)v_{it}. \end{aligned} \quad (\text{A.36})$$

As one would expect, both on the left and right hand sides there are lag polynomials that multiply ε_t and v_{it} . Because the two stochastic process are uncorrelated, the equality must hold independently for the terms that multiply ε_t for those that multiply v_{it} . Taking into account relationship (A.35), equation (A.36) thus defines two fixed points: one for $(L - \lambda)G(L)$ and one for $V(L)$. Differently from the full information case, the fixed point for the aggregate y_t (that defined by the terms multiplying ε_t) also contains elements of the function $V(L)$, more precisely the constant $V(0) - V(\lambda)$. Therefore, in order to solve for $(L - \lambda)G(L)$, we need first to solve for $V(L)$. Taking the fixed point condition for the terms that multiply v_{it} , multiplying both sides by L and rearranging one obtains

$$\begin{aligned} V(L)\phi_x(L) &= \phi_x V(0) - \phi_x \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] L \\ & \quad - \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[\phi_y G(0) + \phi_\theta (A(\lambda) - A(0)) \right] L + \psi_\theta(L)L. \end{aligned} \quad (\text{A.37})$$

where $\phi_x(L) \equiv \phi_x - \psi_x(L)L$. Similarly, the fixed point for $(L - \lambda)G(L)$ is

$$\begin{aligned}
 (L - \lambda)G(L)\Phi(L) = & \\
 & - \phi_x \tilde{\gamma}_y(0)\lambda G(0) - \phi_x (\tilde{\gamma}_\theta(L)A(L) - \tilde{\gamma}_\theta(0)A(0)) + \phi_x (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[X(\lambda) - X(0) - (V(\lambda) - V(0))A(\lambda) \right] L \\
 & + \phi_y \left[\lambda - (1 - \tau(\lambda)) \frac{1-\lambda^2}{(1-\lambda L)} L \right] G(0) - \phi_\theta \left[[A(L) - A(0)] - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} [A(\lambda) - A(0)] L \right] \\
 & + \psi_x(L)\tilde{\gamma}_\theta(L)A(L)L + \psi_\theta(L)A(L)L.
 \end{aligned} \tag{A.38}$$

where we have used (A.35) to substitute for, $X(L) - X(0)$, and, $X(L)$, and, $\Phi(L) \equiv \phi_x(L) + \phi_y - \psi_y(L)L$. The next Lemma will prove very useful.

Lemma A2. $V(\lambda) = \tilde{\gamma}_\theta(\lambda)$.

Proof. Evaluate (A.37) at λ and rearrange to obtain

$$\begin{aligned}
 V(\lambda)\psi_x(\lambda)\lambda = & \\
 & - \phi_x \frac{\tau(\lambda)}{A(\lambda)} [X(\lambda) - X(0)] - \phi_x (1 - \tau(\lambda)) (V(\lambda) - V(0)) - \frac{\tau(\lambda)}{A(\lambda)} [\phi_y G(0)\lambda + \phi_\theta (A(\lambda) - A(0))] + \psi_\theta(\lambda)\lambda.
 \end{aligned} \tag{A.39}$$

Next, evaluate (A.38) at λ and rearrange to obtain

$$\begin{aligned}
 0 = & - \tau(\lambda)\phi_x (X(\lambda) - X(0)) + -\phi_x (1 - \tau(\lambda)) (V(\lambda) - V(0))A(\lambda) - \phi_y \tau(\lambda)G(0)\lambda - \phi_\theta (A(\lambda) - A(0))\tau(\lambda) \\
 & + \psi_x(\lambda)\tilde{\gamma}_\theta(\lambda)A(\lambda)\lambda + \psi_\theta(\lambda)A(\lambda)\lambda.
 \end{aligned} \tag{A.40}$$

Clearly, for (A.39) and (A.40) to hold, assuming $A(\lambda) \neq 0$, $\psi_x(\lambda) \neq 0$ and $\lambda \neq 0$, it must be that $V(\lambda) = \tilde{\gamma}_\theta(\lambda)$. \square

We can now use Lemma A2 to substitute for $V(\lambda)$ in (A.37) and (A.38). It follows that to solve for $(L - \lambda)G(L)$ we just need an expression for $V(0)$, to which we now turn. From Assumption 1, we know that there is a root ζ_V that needs to be removed for $V(L)$ to be stationary. We achieve this by choosing the appropriate constant $V(0)$ so that the numerator on the right hand side of (A.37) vanishes when evaluated at ζ_V ,

$$\begin{aligned}
 & \phi_x V(0) - \phi_x \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda\zeta_V)} [X(\lambda) - X(0) - (\tilde{\gamma}_\theta(\lambda) - V(0))A(\lambda)] \zeta_V \\
 & - \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda\zeta_V)} [\phi_y G(0) + \phi_\theta (A(\lambda) - A(0))] \zeta_V + \psi_\theta(\zeta_V)\zeta_V = 0
 \end{aligned} \tag{A.41}$$

Using (A.35) so substitute for $X(\lambda) - X(0)$, and rearranging one obtain the expression

$$\phi_x V(0)A(\lambda) = m(\lambda)(\phi_x \tilde{\gamma}_y(0) + \phi_y)\lambda G(0) + n(\lambda), \tag{A.42}$$

where

$$m(\lambda) \equiv \frac{\tau(\lambda)(1 - \lambda^2)\zeta_V}{(1 - \lambda\zeta_V)\lambda - \tau(\lambda)(1 - \lambda^2)\zeta_V}, \tag{A.43}$$

and

$$n(\lambda) \equiv \frac{\phi_\theta \tau(\lambda)(1 - \lambda^2)(A(\lambda) - A(0))\zeta_V - \phi_x \tau(\lambda)(1 - \lambda^2)\tilde{\gamma}_\theta(0)A(0)\zeta_V - \psi_\theta(\zeta_V)\zeta_V \lambda A(\lambda)}{(1 - \lambda\zeta_V)\lambda - \tau(\lambda)(1 - \lambda^2)\zeta_V}. \tag{A.44}$$

Next we used (A.42) in (A.38), and we also substitute $X(\lambda) - X(0)$ using (A.35) to get

$$(L - \lambda)G(L) = \frac{-\lambda G(0)(\phi_x \tilde{\gamma}_y(0) + \phi_y)H(L) + J(L)}{\Phi(L)(1 - \lambda L)\lambda}, \tag{A.45}$$

where

$$H(L) = \lambda(1 - \lambda L) - (1 - \tau(\lambda))(1 - \lambda^2)(1 + m(\lambda))L, \tag{A.46}$$

and

$$\begin{aligned}
 J(L) &= (1 - \tau(\lambda))(1 - \lambda^2)[n(\lambda) - \phi_x \tilde{\gamma}_\theta(0)A(0) + \phi_\theta(A(\lambda) - A(0))]L + A(0)(\phi_x \tilde{\gamma}_y(0) + \phi_y)\lambda(1 - \lambda L) \\
 &\quad - [(\phi_x - \psi_x(L)L)\tilde{\gamma}_\theta(L) + \phi_\theta - \psi_\theta(L)L]A(L)\lambda(1 - \lambda L).
 \end{aligned} \tag{A.47}$$

Under Assumption 1, $\Phi(L)$ has a zero inside the unit circle at ζ , which means that we need to choose the constant $G(0)$ so to cancel it. This is achieved by setting

$$-\lambda G(0)(\phi_x \tilde{\gamma}_y(0) + \phi_y)H(\zeta) + J(\zeta) = 0. \tag{A.48}$$

Solving for $G(0)$ and substituting back into (A.45) one gets

$$(L - \lambda)G(L) = \frac{J(L)H(\zeta) - J(\zeta)H(L)}{\Phi(L)(1 - \lambda L)\lambda}. \tag{A.49}$$

Next, recall that we defined

$$\xi_y(L) \equiv A(0)(\phi_x \tilde{\gamma}_y(0) + \phi_y) - [(\phi_x - \psi_x(L)L)\tilde{\gamma}_\theta(L) + \phi_\theta - \psi_\theta(L)L]A(L), \tag{A.50}$$

and letting

$$\tilde{\xi} \equiv n(\lambda) - \phi_x \tilde{\gamma}_\theta(0)A(0) + \phi_\theta(A(\lambda) - A(0)), \tag{A.51}$$

one can show that (A.49) can be written as

$$(L - \lambda)G(L) = \frac{\xi_y(\zeta) - \xi_y(L)}{\Phi(L)} - (1 - \tau(\lambda))(1 - \lambda^2)(\zeta - L)\frac{\tilde{\xi} - (1 + m(\lambda)\xi_y(\zeta))}{H(\zeta)\Phi(L)(1 - \lambda L)}. \tag{A.52}$$

Using the factorization $\Phi(L) = (\zeta - L)\tilde{\Phi}(L)$, and defining

$$\mathcal{A}(\lambda) \equiv \frac{\tilde{\xi} - (1 + m(\lambda)\xi_y(\zeta))}{H(\zeta)}, \tag{A.53}$$

expression (19) follows. Finally, for the solution to be consistent with the information set that we have used to derive it, it must be that the polynomial in (19) vanishes at $L = \lambda$, which corresponds to condition (20) in the Theorem.

The last step of the proof consists in making sure that when the equilibrium coefficients are evaluated using the λ that solves (20), there are no other points at which $Y(L)$ vanishes inside the unit circle. More precisely, it has to be that there is no $\xi \neq \lambda$ that solves

$$\mathcal{Y}(\xi)\tilde{\Phi}(\xi) = (1 - \tau(\lambda))(1 - \lambda^2)\frac{\mathcal{A}(\lambda)}{1 - \lambda\xi}, \tag{A.54}$$

such that $|\xi| \in (-1, 1)$. If this was not the case, then the information conveyed by y_t in equilibrium would be inconsistent with the information used to derive the expectations that we use to determine the fixed point. More precisely, the factorization of $\Gamma(L)$ would be incorrect, as $\Gamma^*(L)$ in (A.24) would still be non-invertible. To see this, suppose that λ is a solution to (20), while ξ is a solution to (A.54), and they are both inside the unit circle. Then, the equilibrium function must have the form $\tilde{G}(L)(L - \lambda)(L - \xi)$, but the factorization above only removes the zero associated with λ . It follows that

$$\Gamma^*(L) = \frac{1}{\sqrt{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}} \begin{bmatrix} A(L)A(\lambda)\sigma_\varepsilon^2 + \sigma_v^2 & \sigma_\varepsilon\sigma_v\frac{1-\lambda L}{L-\lambda}(A(L) - A(\lambda)) \\ A(\lambda)\sigma_\varepsilon^2(L - \lambda)(L - \xi)\tilde{G}(L) & \sigma_\varepsilon\sigma_v\tilde{G}(L)(L - \xi)(1 - \lambda L) \end{bmatrix}, \tag{A.55}$$

whose determinant still vanishes at $L = \xi$, so that $\Gamma^*(L)$ is not the appropriate factorization. In this case one can modify the initial guess and consider $N > 1$ roots inside the unit circle, looking then for a condition like (20) to deliver exactly N solutions. We restrict our attention to $N = 1$ for simplicity and because the full description of the space of REE with confounding dynamics is beyond the scope of this paper, but we hope it is clear that our methods extend to the more general case.

STEP 4: NO INFORMATION FROM THE MODEL The last thing to check to complete the proof is to ensure that there is no information that is transmitted by a clever manipulation of the model conditions – which are part of the information set of the agents – combined with the knowledge of the history of θ_{it} and y_t . For instance, suppose that the market clearing condition (10) is specified so that $\int_0^1 x_{it}\mu(i)di = y_t$, which means that y_t is the aggregate of x_{it} , then this would imply $X(L) = Y(L)$, which would result in $x_{it} - y_t = V(L)v_{it}$. Because rational agents know all this, they know that the difference $x_{it} - y_t$ is just a linear combination of the individual innovations v_{it} . It follows that they could, in principle, back out the realizations of v_{it} 's by inverting $V(L)$. More generally, the link between $X(L)$ and $Y(L)$ due to (10) can be used by rational agents to obtain additional information on the underlying innovations. For this not to happen, if one augments the information set of the agents by $x_{it} - y_t$, the information matrix must still be non-invertible at λ . The following Lemma shows that this is indeed the case for the equilibrium of Theorem 1.

Lemma A3. *In the equilibrium with confounding dynamics of Theorem 1, consider the augmented information matrix $\tilde{\Gamma}(L)$, where*

$$\begin{pmatrix} \theta_{it} \\ y_t \\ x_{it} - y_t \end{pmatrix} = \tilde{\Gamma}(L) \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix} = \begin{bmatrix} A(L) & 1 \\ Y(L) & 0 \\ X(L) - Y(L) & V(L) \end{bmatrix} \begin{pmatrix} \varepsilon_t \\ v_{it} \end{pmatrix}. \quad (\text{A.56})$$

The 2-by-2 minors of $\tilde{\Gamma}(L)$ all vanish at λ .

Proof. Matrix $\tilde{\Gamma}(L)$ has three minors, whose determinants are, respectively, $Y(L)$, $Y(L)V(L)$, and, $A(L)V(L) - (X(L) - Y(L))$. The first two minors clearly vanish at λ since, by construction, $Y(\lambda) = 0$. For the third minor, use (A.35) to write

$$A(L)V(L) - (X(L) - Y(L)) = A(L)V(L) - \tilde{\gamma}_y(L)Y(L) - \tilde{\gamma}_\theta(L)A(L) + Y(L). \quad (\text{A.57})$$

We thus need to show that

$$A(\lambda)V(\lambda) = \tilde{\gamma}_\theta(\lambda)A(\lambda), \quad (\text{A.58})$$

but this follows immediately from Lemma A2. \square

A.3 DERIVATION OF FULL INFORMATION SOLUTION OF RBC MODEL In terms of the notation we used in Appendix A.1, $\mathcal{Y}(L)$ would correspond to $\mathcal{R}(L)$ in the application of Section 5, and $\mathcal{X}(L)$ to $\mathcal{K}(L)$. We first note that here $\gamma_x(L) = (1 - \alpha)L$, which means $\gamma_x(0) = 0$, we are thus forced to take the alternative route described in Appendix A.1 and solve first for $\mathcal{K}(L)$ and leave the solution to $\mathcal{R}(L)$ as a straightforward corollary. Under full information we know that $\mathbb{E}_{it}(k_{it+2}) = [\mathcal{K}(L) - \mathcal{K}(0)]L^{-1}\varepsilon_t + [\mathcal{V}(L) - \mathcal{V}(0)]L^{-1}v_{it}$, $\mathbb{E}_{it}(k_{it+1}) = k_{t+1} = \mathcal{K}(L)\varepsilon_t$, and $\mathbb{E}_{it}(a_{it+1}) = \mathbb{E}_{it}(a_{it+1}) = [A(L) - A(0)]L^{-1}\varepsilon_t$. Substituting (30) into (31), using the above expressions for the expectations, aggregating over agents, multiplying both sides by L , and rearranging, one obtains the fixed point condition

$$\mathcal{K}(L) = \frac{\alpha\beta\mathcal{K}(0) + (1 - \eta(1 - \alpha\beta) - L)A(L) - A(0)}{\alpha(\zeta - L)(\beta/\zeta - L)}. \quad (\text{A.59})$$

To ensure stationarity we choose $\mathcal{K}(0) = \frac{1}{\alpha\beta}(A(0) - (1 - \eta(1 - \alpha\beta) - \zeta)A(\zeta))$. Next substitute this expression in (A.59), and specify $A(L) = \frac{1}{1 - \rho L}$. By construction the denominator polynomial contains the factor $(\zeta - L)$, which can be easily isolated and simplified with the same factor at the denominator, so to finally obtain

$$\mathcal{K}(L) = \frac{\frac{\zeta}{\alpha\beta} \left(\frac{1 - (1 - \eta(1 - \alpha\beta))\rho}{1 - \rho\zeta} \right)}{(1 - \rho L)(1 - \frac{\zeta}{\beta}L)}. \quad (\text{A.60})$$

Evaluating the characteristic polynomial (34) at 1 one can show that, $\alpha(\zeta - 1)(\beta/\zeta - 1)/(1 - \alpha) = \eta(1 - \alpha\beta)$. Adding, $\zeta - 1$, on both sides and rearranging one can show that, $\alpha(1 - \zeta)(\alpha\beta/\zeta - 1)/(1 - \alpha) = \eta(1 - \alpha\beta) - 1 + \zeta$. Now take

the term $\frac{1-(1-\eta(1-\alpha\beta))\rho}{1-\rho\zeta}$, add and subtract $\rho\zeta$ at the numerator, to obtain

$$\mathcal{K}(L) = \frac{\frac{\zeta}{\alpha\beta}(1+\kappa)}{(1-\rho L)(1-\frac{\zeta}{\beta}L)}. \quad (\text{A.61})$$

where $\kappa \equiv \frac{\rho(1-\zeta)(\alpha\beta/\zeta-1)}{(1-\rho\zeta)(1-\alpha)}$. It can be showed that $\kappa = 0$ for $\eta = 1$, which corresponds to the case of logarithmic preferences, and $\kappa > 0$ (resp. < 0) when $\eta < 1$ (resp. > 1). The expression for $\mathcal{R}(L)$ can be obtained using the relationship, $\mathcal{R}(L) = A(L) - (1-\alpha)\mathcal{K}(L)L$.

A.4 PROOF OF COROLLARY 1 The proof of the corollary is a straightforward application of the following lemma.

Lemma A4. *Consider the Real Business Cycle model (30)-(31). Let the information sets be specified as in (37). There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form, $k_{t+1} = K(L)\varepsilon_t$, and $r_t = R(L)\varepsilon_t$, with*

$$K(L) = \mathcal{K}(L) - (1-\tau(\lambda))(1-\lambda^2)\frac{\mathcal{A}_k(\lambda)}{(1-\lambda L)(\tilde{\zeta}-L)}, \quad (\text{A.62})$$

and, $R(L) = A(L) - (1-\alpha)K(L)L$, if there exists a $\lambda \in (-1, 1)$, that solves

$$\mathcal{R}(\lambda)(\lambda - \tilde{\zeta}) = (1-\alpha)(1-\tau(\lambda))\mathcal{A}_k(\lambda)\lambda, \quad (\text{A.63})$$

where $\mathcal{K}(L)$ and $\mathcal{R}(L)$ are the full information solutions, $\tau(\lambda) \equiv \frac{A(\lambda)^2\sigma_\varepsilon^2}{A(\lambda)^2\sigma_\varepsilon^2 + \sigma_v^2}$, $\mathcal{A}_k(\lambda)$ is a function of λ that depends only on exogenous parameters, and $R(L)$ has a zero inside the unit circle equal to λ .

Proof. The proof follows the same steps as that of Theorem 1, with the difference that we solve for $X(L)$ first – $K(L)$ in the application. Recall that

$$\phi_x = \alpha\beta, \quad \phi_y = 1 - \alpha\beta, \quad \phi_\theta = 1, \quad \psi_x(L) = \alpha(1 + \beta) - \alpha L, \quad \psi_y(L) = 0, \quad \psi_\theta(L) = -1.$$

and

$$\gamma_x(L) = (1-\alpha)L, \quad \gamma_y(L) = 1, \quad \gamma_\theta(L) = -1.$$

Note that, although the notation adopted in the model has the two variables having different time subscripts, r_t and k_{t+1} , they are both pre-determined at time t , and so they are both functions of possibly the infinite history of ε_t up to time t . Since we are looking for an equilibrium with confounding dynamics, we operationalize the condition $R(\lambda) = 0$ by conjecturing

$$R(L) = (L - \lambda)G(L), \quad (\text{A.64})$$

where $G(L)$ has no zeros inside the unit circle. Because in equilibrium $R(L) = A(L) - (1-\alpha)K(L)L$, the conjecture immediately implies

$$A(\lambda) = (1-\alpha)K(\lambda)\lambda, \quad (\text{A.65})$$

a relationship that will be useful in what follows. One important remark on (A.65) is that it implies $\lambda \neq 0$. In fact, evaluating the expression at $\lambda = 0$, provided that $K(0)$ is well defined, which must be the case in the solution we want to characterize, gives $A(0) = 0$, which never holds by assumption. Hence, the statement of the Proposition requires $|\lambda| \in (0, 1)$. The information set takes the form of (A.22), where $x_{it} = a_{it}$ and $y_t = r_t$, so that $\mathbb{E}_{it}(a_{t+1})$ and $\mathbb{E}_{it}(r_{t+1})$ are provided by (A.28) and (A.29), respectively. For the term $\mathbb{E}_{it}(k_{it+2})$ things require some extra steps. We work under the conjecture that

$$k_{it+1} = K(L)\varepsilon_t + V(L)v_{it}, \quad (\text{A.66})$$

Next, we evaluate the variance-covariance generating function between the information set and k_{it+1} , which is

$$g_{k_i, (a_i, r)}(z) = \begin{bmatrix} K(z)A(z^{-1})\sigma_\varepsilon^2 + V(z)\sigma_v^2 & K(z)(z^{-1} - \lambda)G(z^{-1})\sigma_\varepsilon^2 \end{bmatrix}. \quad (\text{A.67})$$

We then use this expression, together with the canonical factorization $\Gamma^*(z)$ in (A.24) in the Wiener-Kolmogorov

formula (C.31), and following steps similar to (A.32) and (A.33) to finally get

$$\begin{aligned} \mathbb{E}_{it}(k_{it+2}) &= L^{-1} \left[K(L) - K(0) \right] \varepsilon_t - (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[K(0) - K(\lambda) - (V(0) - V(\lambda))A(\lambda) \right] \varepsilon_t \\ &\quad + L^{-1} \left[V(L) - V(0) \right] v_{it} - \tau(\lambda) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[\frac{K(0)-K(\lambda)}{A(\lambda)} + (V(0) - V(\lambda)) \right] v_{it}. \end{aligned} \quad (\text{A.68})$$

We can now use the expressions for the expectational terms to obtain a fixed point condition similar to (A.36),

$$\begin{aligned} \alpha(1 + \beta)K(L)\varepsilon_t + \alpha(1 + \beta)V(L)v_{it} &= \alpha\beta L^{-1} \left[K(L) - K(0) \right] \varepsilon_t + \alpha\beta L^{-1} \left[V(L) - V(0) \right] v_{it} \\ &\quad - \alpha\beta(1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[K(\lambda) - K(0) - (V(\lambda) - V(0))A(\lambda) \right] \varepsilon_t + \alpha\beta \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} \left[K(\lambda) - K(0) - (V(\lambda) - V(0))A(\lambda) \right] v_{it} \\ &\quad + \alpha K(L)L\varepsilon_t + \alpha V(L)Lv_{it} + A(L)\varepsilon_t + v_{it} - [A(L) - A(0)]L^{-1}\varepsilon_t + (1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} [A(\lambda) - A(0)]\varepsilon_t \\ &\quad - \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} [A(\lambda) - A(0)]v_{it} + (1 - \alpha\beta) \left[A(L) - (1 - \alpha)K(L)L - A(0) \right] L^{-1}\varepsilon_t \\ &\quad + (1 - \alpha\beta)(1 - \tau(\lambda)) \frac{1-\lambda^2}{\lambda(1-\lambda L)} A(0)\varepsilon_t - (1 - \alpha\beta) \frac{\tau(\lambda)}{A(\lambda)} \frac{1-\lambda^2}{\lambda(1-\lambda L)} A(0)v_{it}, \end{aligned} \quad (\text{A.69})$$

where we have used $(L - \lambda)G(L) = A(L) - (1 - \alpha)K(L)L$, and thus $-\lambda G(0) = A(0)$, to substitute for terms related to $G(L)$. The fixed point equation contains only terms related to the endogenous polynomials $V(L)$ and $K(L)$, and one can proceed to solve for the fixed point as in the proof of Theorem 1. In particular, using the same steps as in Lemma A2, one can show that $A(\lambda)V(\lambda) = K(\lambda)$, and, in addition, we know that (A.65) holds, so we can set $K(\lambda) = \frac{A(\lambda)}{\lambda(1-\alpha)}$. The uniqueness of a stationary solution under Assumption 1 and condition (34), is once again obtained by the appropriate choice of $V(0)$ and $K(0)$. In the end, the expression for $\mathcal{A}_k(\lambda)$, analogue to the constant $\mathcal{A}(\lambda)$ in Theorem 1, can be simplified to

$$\mathcal{A}_k(\lambda) = \frac{[(1-\lambda\beta)\lambda - \tau(\lambda)(1-\lambda^2)\beta] \left[(1-\beta)A(\lambda) + \beta(\eta(1-\beta\alpha) - 1) \right] (1-\tau(\lambda)) (1-\lambda^2)A(0) + (1-\lambda\beta)\lambda \left[\eta(1-\beta\alpha) - 1 + \zeta \right] A(\zeta)}{\lambda\alpha(1+\beta) \left[(1-\lambda\beta)(\zeta - \lambda) - \tau(\lambda)(1-\lambda^2)(\zeta - \beta) \right]} \quad (\text{A.70})$$

The condition for the existence of one $|\lambda| \in (0, 1)$ follows from using $K(L)$ to write $R(L)$ and then imposing $R(\lambda) = 0$. The same argument that we have used in the proof of Theorem 1 to argue that when the equilibrium coefficients are evaluated using the λ that solves $R(\lambda) = 0$, there must be no other points at which $R(L)$ vanishes inside the unit circle, applies here as well. This completes the proof. \square

The proof of Corollary 1 consists in plugging $A(L) = \frac{1}{1-\rho L}$ into the above expressions and rearranging terms when possible.

APPENDIX FOR ONLINE PUBLICATION

B ADDITIONAL RESULTS AND DERIVATIONS

B.1 DERIVATION OF PREDICTIONS IN SECTION 2 The critical step is the application of the annihilating operator function

$$\left[\sigma_\varepsilon^2 (L^{-1} - \lambda) [(1 - \lambda L^{-1}) \sigma_\varepsilon]^{-1} \right]_+ = \sigma_\varepsilon \left[\frac{L^{-1} - \lambda}{1 - \lambda L^{-1}} \right]_+. \quad (\text{B.1})$$

Multiply both numerator and denominator of the right hand side by L to get

$$\sigma_\varepsilon \left[\frac{L^{-1} - \lambda}{1 - \lambda L^{-1}} \right]_+ = \sigma_\varepsilon \left[\frac{1 - \lambda L}{L - \lambda} \right]_+. \quad (\text{B.2})$$

The argument is now a regular function that has an isolated singularity at $L = \lambda$. The principal part of the Laurent series expansion is determined as

$$\lim_{z \rightarrow \lambda} (z - \lambda) \frac{1 - \lambda z}{z - \lambda} = (1 - \lambda^2). \quad (\text{B.3})$$

Using Lemma C2 and after some algebra we have

$$\left[\frac{1 - \lambda L}{L - \lambda} \right]_+ = \frac{1 - \lambda L}{L - \lambda} - \frac{1 - \lambda^2}{L - \lambda} = \frac{\lambda(L - \lambda)}{L - \lambda} = -\lambda. \quad (\text{B.4})$$

In Section B.4 we show that (B.4) generalizes to the case of N singularities.

B.2 EQUIVALENCE IN SIGNAL EXTRACTION We need to show that the representations (1) and (5) are equivalent in terms of unconditional forecast error variance

$$\mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t | s^t))^2 \right] = \mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t | z^t))^2 \right] \quad (\text{B.5})$$

when $\lambda^2 = \tau$, where $\tau = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\eta^2}$. The optimal forecast $\mathcal{P}[\varepsilon_t | z^t]$ is given by weighting z_t according to the relative variance of ε , $\mathcal{P}(\varepsilon_t | z^t) = \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_\eta^2} \right) z_t = \tau z_t$ and therefore,

$$\mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t | z^t))^2 \right] = \frac{\sigma_\varepsilon^2 \sigma_\eta^2}{\sigma_\varepsilon^2 + \sigma_\eta^2} = (1 - \tau) \sigma_\varepsilon^2. \quad (\text{B.6})$$

From (3) we know that when $|\lambda| < 1$, $\mathcal{P}(\varepsilon_t | s^t) = -\lambda \left(\frac{L - \lambda}{1 - \lambda L} \right) \varepsilon_t$. It follows that

$$\mathbb{E} \left[(\varepsilon_t - \mathcal{P}(\varepsilon_t | s^t))^2 \right] = \mathbb{E} \left[\left(\varepsilon_t + \lambda \left(\frac{L - \lambda}{1 - \lambda L} \right) \varepsilon_t \right)^2 \right] = (1 - \lambda^2)^2 \mathbb{E} \left[\left(\frac{1}{1 - \lambda L} \varepsilon_t \right)^2 \right] = (1 - \lambda^2)^2 \sigma_\varepsilon^2. \quad (\text{B.7})$$

It follows that the mean-squared forecast errors (B.6) and (B.7) are equal when $\lambda^2 = \tau$.

B.3 CONFOUNDING DYNAMICS WITH $m > n$ Let the signal structure be specified as in the text

$$s_t = \Gamma(L) u_t, \quad (\text{B.8})$$

where $\Gamma(L)$ is $n \times m$. The objective of this section is to provide a formal definition of confounding dynamics that applies when $m \geq n$, and then provide two examples for the case of $m = 2$ and $n = 1$. Denote the variance covariance matrix of the signal vector by $g_{ss}(z)$. The matrix $g_{ss}(z)$ is an $n \times n$ positive-definite matrix of rank $r \leq n$ for $|z| = 1$, and with rational elements. The final goal here is to find an appropriate factorization of $g_{ss}(z)$ that can be used in the Wiener-Kolmogorov prediction formula. To that end, we follow Rozanov (1967), pages 44-47, and we divide the factorization in two steps. First, we perform a factorization that delivers a function $\tilde{\Gamma}(z)$ which is $n \times r$, and that has

rational elements and it is analytic inside the unit circle. Next, we check whether the function $\tilde{\Gamma}(z)$ has rank r for z inside the unit circle. If not, then it means that there exists one or more point in which all the minors of order r of $\tilde{\Gamma}(z)$ vanish. For our purposes, this means that the original signal structure $\Gamma(L)$ not only is non-invertible because $m > n$, but also because of confounding dynamics.

We begin here by stating the existence of the function $\tilde{\Gamma}(z)$ as a Lemma.

Lemma B1. *A positive definite matrix function $g_{ss}(z)$ of dimension $n \times n$, and of rank $r \leq n$, with elements which are rational functions of z , can be represented in the form*

$$g_{ss}(z) = \tilde{\Gamma}(z)\tilde{\Gamma}(z^{-1})^\top, \quad (\text{B.9})$$

where $\tilde{\Gamma}(z)$ is $n \times r$, and the elements in $\tilde{\Gamma}(z)$ are rational with respect to z and analytic inside the unit circle.

Proof. See Rozanov (1967), pages 44-46. □

The general definition of confounding dynamics immediately follows.

Definition GCD. *Let s_t be specified as in (B.8), with $m \geq n$, $\text{rank}(g_{ss}(z)) = r \leq n$ for $|z| = 1$, and let $\tilde{\Gamma}(z)$ be defined as in Lemma B1. The s_t process is said to display confounding dynamics if there exists some λ with $|\lambda| < 1$, such that $\text{rank}(\tilde{\Gamma}(\lambda)) < r$.*

Note first that for $m = n$, and $r = n$, one has that $\tilde{\Gamma}(L) = \Gamma(L)$, and the above definition corresponds to the equilibrium with confounding dynamics as defined in the main text. For the case $m > n$, one necessarily has that $r < m$, which is a formal way to express the fact that the initial signal system is not able to perfectly reveal the history of u_t . Matrix $\tilde{\Gamma}(z)$ of Lemma B1 performs a linear combination of the m elements in u_t into at most r orthogonal components, so to have a representation of the signals s_t that can be used for optimal prediction. However, such linear combination might come short of providing r orthogonal components when the rank of $\tilde{\Gamma}(z)$ is less than r for some z inside the unit circle. When that happens, it means that the r orthogonal components are combined in such a way that their information is confounded into fewer than r orthogonal components, i.e. confounding dynamics are present. Once $\tilde{\Gamma}(z)$ is obtained and the set of λ 's from the definition above identified, the derivation of the canonical factorization $\Gamma^*(z)$ follows the steps outlined in Appendix C.3. The canonical factorization $\Gamma^*(z)$ finally returns a representation of s_t into r orthogonal components, with variances that provide the least-squares prediction.

We present two examples of signal systems with $n = 1$ and $m = 2$, which contains confounding dynamics.

EXAMPLE 1. Consider a process s_t specified as

$$s_t = (L - \lambda)(\varepsilon_t + v_t), \quad (\text{B.10})$$

where $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon)$, $v_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v)$, and $\lambda < 0$. Here $\Gamma(L)$ is the 1×2 matrix

$$\Gamma(L) = \left((L - \lambda)\sigma_\varepsilon \quad (L - \lambda)\sigma_v \right), \quad (\text{B.11})$$

and the variance-covariance generating function is,

$$g_{ss}(z) = (z - \lambda)(z^{-1} - \lambda)(\sigma_\varepsilon^2 + \sigma_v^2). \quad (\text{B.12})$$

Here $r = 1$ since $g_{ss}(z)$ is non-zero when evaluated at $|z| = 1$. Applying Lemma B1, one has that $\tilde{\Gamma}(L) = \sqrt{\sigma_\varepsilon^2 + \sigma_v^2}(L - \lambda)$. Application of Definition B.3 informs that the 1×2 process s_t has confounding dynamics provided that $\lambda \in (-1, 1)$. Using the procedure in Appendix C.3 one can show that the canonical factorization here is $\Gamma^*(L) = \sqrt{\sigma_\varepsilon^2 + \sigma_v^2}(1 - \lambda L)$. Plugging this into the Wiener-Kolmogorov formula for the mean-squared error minimizing prediction $\mathcal{P}(\varepsilon_t | s^t)$ one obtains,

$$\mathcal{P}(\varepsilon_t | s^t) = -\tau \frac{\lambda}{1 - \lambda L} s_t, \quad (\text{B.13})$$

where $\tau \equiv \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_v^2}$, as usual. Equation (B.13) clearly shows how the two sources of non-invertibility combine: the dimension $m > n$ turns into the signal-to-noise coefficient τ , while confounding dynamics turn into the dynamic operator $-\frac{\lambda}{1-\lambda L}$. In higher dimensional system the combination takes a substantially more convoluted form.

EXAMPLE 2. Consider a process s_t specified as

$$s_t = -\lambda\varepsilon_t + \varepsilon_{t-1} + v_t, \quad (\text{B.14})$$

where, once again, $\varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_\varepsilon)$, $v_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma_v)$, and $\lambda < 0$. Here $\Gamma(L)$ is the 1×2 rectangular matrix

$$\Gamma(L) = \begin{pmatrix} (L - \lambda)\sigma_\varepsilon & \sigma_v \end{pmatrix}. \quad (\text{B.15})$$

The variance-covariance generating function is

$$g_{ss}(z) = \sigma_\varepsilon^2(z - \lambda)(z^{-1} - \lambda) + \sigma_v^2. \quad (\text{B.16})$$

According to Lemma B1, we have that $\tilde{\Gamma}(L) = (1 - \hat{\lambda}L)\sigma_w$, where $\hat{\lambda}$ is a solution to the quadratic equation

$$\lambda\sigma_\varepsilon^2 - ((1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2)\hat{\lambda} + \lambda\sigma_\varepsilon^2\hat{\lambda}^2, \quad (\text{B.17})$$

and σ_w^2 is determined by $\sigma_w^2\hat{\lambda} = \lambda\sigma_\varepsilon^2$. To see this note that $g_{ss}(z)$ can be rewritten as

$$g_{ss}(z) = \sigma_\varepsilon^2(1 - z\lambda)(1 - z^{-1}\lambda) + \sigma_v^2, \quad (\text{B.18})$$

and we are looking for the factorization

$$\sigma_\varepsilon^2(1 - z\lambda)(1 - z^{-1}\lambda) + \sigma_v^2 = \sigma_w^2(1 - z\hat{\lambda})(1 - z^{-1}\hat{\lambda}). \quad (\text{B.19})$$

Note that (B.19) can be written as

$$(1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2 - \lambda\sigma_\varepsilon^2z^{-1} - \lambda\sigma_\varepsilon^2z = \sigma_w^2(1 + \hat{\lambda}^2) - \sigma_w^2\hat{\lambda}z^{-1} - \sigma_w^2\hat{\lambda}z. \quad (\text{B.20})$$

Matching coefficients we get two conditions in two unknowns, namely

$$\sigma_w^2\hat{\lambda} = \lambda\sigma_\varepsilon^2, \quad (\text{B.21})$$

and

$$\sigma_w^2(1 + \hat{\lambda}^2) = (1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2. \quad (\text{B.22})$$

Using (B.21) to substitute for σ_w in (B.22) one gets the quadratic equation (B.17).¹⁵ Note that our candidate $\tilde{\Gamma}(L)$ satisfies the requirements of Lemma B1 independently of which root $\hat{\lambda}$ is chosen. However, the quadratic form (B.17) is such that there is always one root inside and one root outside the unit circle. The roots are

$$\hat{\lambda} = \frac{(1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2 \pm \sqrt{((1 + \lambda^2)\sigma_\varepsilon^2 + \sigma_v^2)^2 - 4\lambda^2\sigma_\varepsilon^4}}{2\lambda\sigma_\varepsilon^2}. \quad (\text{B.23})$$

Taking the limit of (B.23) for $\sigma_v \rightarrow 0$, which corresponds to the case presented in Section 2, one has that

$$\lim_{\sigma_v \rightarrow 0} \hat{\lambda}_+ = \lambda^{-1}, \quad \lim_{\sigma_v \rightarrow 0} \hat{\lambda}_- = \lambda, \quad (\text{B.24})$$

where the first root is the one associated with the “+” sign for the discriminant term, and the second root the one associated with the “−” sign. Suppose that we set $\hat{\lambda} = \hat{\lambda}_+$ in $\tilde{\Gamma}(L)$, then according to our Definition GDC the process s_t has confounding dynamics whenever $|\hat{\lambda}_+| > 1$, which is always the case when $|\lambda| < 1$. The canonical factorization in this case is $\Gamma^*(L) = (1 - \lambda^*L)\sigma_w$, where $\lambda^* = \hat{\lambda}_-$, and σ_w^2 solves (B.21). Application of the Wiener-Kolmogorov

¹⁵This procedure is similar to the one outlined in Sargent (1987), Chapter XI, pages 300-302.

formula then leads to

$$P(\varepsilon_t | s^t) = -\frac{\lambda^*}{1 - \lambda^* L} s_t = -\lambda^* \left(\frac{L - \lambda}{1 - \lambda^* L} \right) \varepsilon_t - \frac{\lambda^*}{1 - \lambda^* L} v_t. \quad (\text{B.25})$$

Expression (B.25) shows that the confounding dynamics hallmark is retained even in presence of exogenous noise. The factor that multiplies ε_t in (B.25) has the same format as the factor in (3), with the only difference that the impact is now scaled by λ^* , and the autoregressive root is also λ^* . One important difference in the rectangular case is that the additional noise term also appears in the prediction function, which is represented by the term that multiplies v_t in (B.25). The noise term has a persistent effect on the prediction, with the same autoregressive root as the ε_t term.

Definition CGD and Examples 1 and 2 show that confounding dynamics can be present in any type of signal structure. The signal matrix structure with $m = n$ that we employ in the main text is analytically convenient, and, at the same time, expressions (B.13) and (B.25) suggest that it is without loss of generality, in so far as the purpose is to qualitatively characterize confounding dynamics.

B.4 N NON-INVERTIBLE ROOTS The following proposition describes the prediction formula for the innovations of a process with N non-invertible roots.

Proposition B1. *Let*

$$s_t = \prod_{i=1}^N (L - \lambda_i) \varepsilon_t. \quad (\text{B.26})$$

with $|\lambda_i| < 1$, for $i = 1, \dots, N$. The least squares prediction $\mathcal{P}(\varepsilon_t | s^t)$ is given by

$$P(\varepsilon_t | s^t) = \prod_{i=1}^N \lambda_i \frac{L - \lambda_i}{1 - \lambda_i L} \varepsilon_t. \quad (\text{B.27})$$

Proof. The first step in the proof is to figure out the canonical factorization of $g_{ss}(z)$ when s_t is as in (B.26). Rozanov (1967) method applies directly here so that

$$\Gamma^*(z) = \sigma_\varepsilon \prod_{i=1}^N (1 - \lambda_i L). \quad (\text{B.28})$$

The application of the Wiener-Kolmogorov prediction formula results in the following

$$\Pi_0(z) = \left[\sigma_\varepsilon^2 \prod_{i=1}^N (\lambda_i - z^{-1}) \left[\sigma_\varepsilon \prod_{i=1}^N (1 - \lambda_i z^{-1}) \right]^{-1} \right]_+ \left(\prod_{i=1}^N (1 - \lambda_i L) \right)^{-1}. \quad (\text{B.29})$$

The next Lemma is useful in solving for the annihilating operator.

Lemma B2.

$$\left[\prod_{i=1}^N \frac{\lambda_i - z^{-1}}{1 - \lambda_i z^{-1}} \right]_+ = \prod_{i=1}^N \lambda_i. \quad (\text{B.30})$$

The proof of the Lemma is by induction, repeatedly using Lemma C2 to obtain a solution for $N = 1, 2, 3, \dots$ \square

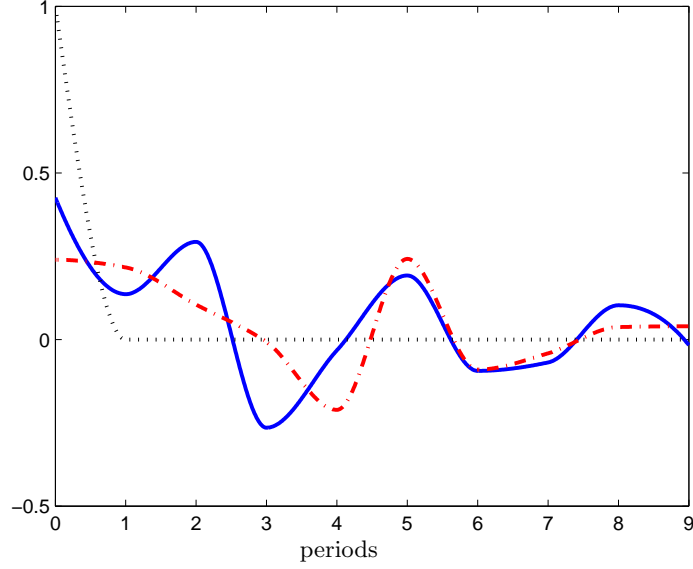
Application of the Lemma leads to

$$\Pi_0(z) = \prod_{i=1}^N \frac{\lambda_i}{1 - \lambda_i z}. \quad (\text{B.31})$$

The final result of the proposition then immediately follows.

Figure 3 shows the impulse response of the prediction formula for $N = 2$ and $N = 3$. As one can see, the mechanical alternation of over-reaction and under-reaction typical of the $N = 1$ case is lost here. In fact, longer and asymmetric cycles become clearly possible.

Figure 3: Prediction Impulse Response



Impulse response of prediction formula $P(\varepsilon_t | s^t)$, for $N = 2$ (solid blue line), and $N = 3$ (dashed red line). The non-invertible roots are $\lambda_1 = -0.3 + 0.75i$, $\lambda_2 = -0.3 - 0.75i$, and $\lambda_3 = -0.75$.

B.5 WEIGHTED SUM OF EXPECTATIONS REPRESENTATION We consider equation (7) with $\phi_\theta = 0$, $\varphi_x(L) = 1$, $\varphi_y(L) = 0$, $\varphi_\theta(L) = -1$, and equation (10) with $\gamma_x(L) = \gamma_y(L)$, and $\gamma_\theta(L) = 0$, so that

$$x_{it} = \phi_x \mathbb{E}_{it}(x_{it+1}) + \phi_y \mathbb{E}_{it}(y_{t+1}) + \theta_{it} \quad (\text{B.32})$$

and

$$y_t = \int_0^1 x_{it} \mu(i) di. \quad (\text{B.33})$$

Under the form we have assumed for θ_{it} , we always have that for $j \geq 1$,

$$\mathbb{E}_{it+j}(\theta_{it+j+1}) = \mathbb{E}_{it+j}(\theta_{t+j+1}), \quad (\text{B.34})$$

a property that will keep the notation below manageable. To initiate the iterative substitution, take (B.32) one period forward so that

$$x_{it+1} = \phi_x \mathbb{E}_{it+1}(x_{it+2}) + \phi_y \mathbb{E}_{it+1}(y_{t+2}) + \theta_{it+1}. \quad (\text{B.35})$$

Next aggregate (B.35) and apply (B.33) to get

$$y_{t+1} = \phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2}) + \theta_{t+1}. \quad (\text{B.36})$$

Now we can use (B.35) and (B.36) to substitute for x_{it+1} and y_{t+1} in (B.32) to get

$$\begin{aligned} x_{it} &= \phi_x \mathbb{E}_{it} [\phi_x \mathbb{E}_{it+1}(x_{it+2}) + \phi_y \mathbb{E}_{it+1}(y_{t+2}) + \theta_{it+1}] + \phi_y \mathbb{E}_{it} [\phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2}) + \theta_{t+1}] + \theta_{it} \\ &= \phi_x \phi_x \mathbb{E}_{it}(x_{it+2}) + \phi_x \phi_y \mathbb{E}_{it}(y_{t+2}) + \phi_y \mathbb{E}_{it} [\phi_x \bar{\mathbb{E}}_{t+1}(x_{it+2}) + \phi_y \bar{\mathbb{E}}_{t+1}(y_{t+2})] + (\phi_x + \phi_y) \mathbb{E}_{it}(\theta_{t+1}) + \theta_{it} \end{aligned} \quad (\text{B.37})$$

where, in the second line, we applied the Law of Iterated Expectations (LIE), when possible, and we made use of (B.34). Next we want to substitute for x_{it+2} and y_{t+2} , so we carry (B.32) two periods forward so that

$$x_{it+2} = \phi_x \mathbb{E}_{it+2}(x_{it+3}) + \phi_y \mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}. \quad (\text{B.38})$$

aggregate (B.38), and apply (B.33) to get

$$y_{t+2} = \phi_x \bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y \bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}. \quad (\text{B.39})$$

Now we can use (B.38) and (B.39) to substitute for x_{it+2} and y_{t+3} in (B.37) to get

$$\begin{aligned} x_{it} &= \phi_x \phi_x \mathbb{E}_{it} [\phi_x \mathbb{E}_{it+2}(x_{it+3}) + \phi_y \mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}] + \phi_x \phi_y \mathbb{E}_{it} [\phi_x \bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y \bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}] \\ &+ \phi_y \mathbb{E}_{it} \{ \phi_x \bar{\mathbb{E}}_{t+1} [\phi_x \mathbb{E}_{it+2}(x_{it+3}) + \phi_y \mathbb{E}_{it+2}(y_{t+3}) + \theta_{it+2}] + \phi_y \bar{\mathbb{E}}_{t+1} [\phi_x \bar{\mathbb{E}}_{t+2}(x_{it+3}) + \phi_y \bar{\mathbb{E}}_{t+2}(y_{t+3}) + \theta_{t+2}] \} \\ &+ (\phi_x + \phi_y) \mathbb{E}_{it}(\theta_{t+1}) + \theta_{it} \\ &= \phi_x \phi_x \phi_x \mathbb{E}_{it}(x_{it+3}) + \phi_x \phi_x \phi_y \mathbb{E}_{it}(y_{t+3}) + \phi_x \phi_y \phi_x \mathbb{E}_{it}[\bar{\mathbb{E}}_{t+2}(x_{it+3})] + \phi_x \phi_y \phi_y \mathbb{E}_{it}[\bar{\mathbb{E}}_{t+2}(y_{t+3})] \\ &+ \phi_y \phi_x \phi_x \mathbb{E}_{it} \{ \bar{\mathbb{E}}_{t+1} [\mathbb{E}_{it+2}(x_{it+3})] \} + \phi_y \phi_x \phi_y \mathbb{E}_{it} \{ \bar{\mathbb{E}}_{t+1} [\mathbb{E}_{it+2}(y_{t+3})] \} \\ &+ \phi_y \phi_y \phi_x \mathbb{E}_{it} \{ \bar{\mathbb{E}}_{t+1} [\bar{\mathbb{E}}_{t+2}(x_{it+3})] \} + \phi_y \phi_y \phi_y \mathbb{E}_{it} \{ \bar{\mathbb{E}}_{t+1} [\bar{\mathbb{E}}_{t+2}(y_{t+3})] \} \\ &+ \phi_x \phi_x \mathbb{E}_{it}(\theta_{t+2}) + \phi_x \phi_y \mathbb{E}_{it}(\theta_{t+2}) + \phi_y \phi_x \mathbb{E}_{it}[\bar{\mathbb{E}}_{t+1}(\theta_{t+2})] + \phi_y \phi_y \mathbb{E}_{it}[\bar{\mathbb{E}}_{t+1}(\theta_{t+2})] + (\phi_x + \phi_y) \mathbb{E}_{it}(\theta_{t+1}) + \theta_{it}. \end{aligned} \quad (\text{B.40})$$

Proceeding in such manner up to some arbitrary time J , one ends up getting a weighted sum of expectations of different orders about θ_{t+j} , for $j = 1, \dots, J+1$, while the remaining endogenous variables x_{it+J+1} and y_{t+J+1} multiply coefficients that tend to zero as. Letting $J \rightarrow \infty$ and aggregating over agents, one finally obtains

$$y_t = \sum_{j=1}^{\infty} \mathcal{P}_{\phi_x}^j \left[(\phi_x + \phi_y)^j \bar{\mathbb{E}}^j(\theta_{t+j}) \right] \quad (\text{B.42})$$

where $\bar{\mathbb{E}}^j(\theta_{t+j})$ stands for the j^{th} order average expectation of θ_{t+j} , and, for notational convenience, we let $\bar{\mathbb{E}}^0(\theta_t) = \theta_t$. The way the operator $\mathcal{P}_{\phi_x}^j$ works is visible in the first four terms of the last line of (B.41) where $j = 2$. In the first and second term, ϕ_x appears as the first coefficient, which results in the expectation being of the first order. In the last two terms, ϕ_y appears as the first coefficient, which results in the expectation being of the second order. In subsequent substitution the pattern that results in the reduction of some of the higher order compounding is quite complex, as the combination of relative positions of ϕ_x and ϕ_y in the coefficients grows at the power of 2^n , so we omit it here. For instance, if one considers the expression at the 3rd iteration, i.e. for $J = 3$, then one has

$$\begin{aligned} y_t &= [\text{terms with } x_{it+4}, y_{t+4}] + (\phi_x \phi_x \phi_x + \phi_x \phi_x \phi_y) \bar{\mathbb{E}}_t(\theta_{t+3}) + (\phi_x \phi_y \phi_x + \phi_x \phi_y \phi_y) \bar{\mathbb{E}}_t[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})] \\ &+ (\phi_y \phi_y \phi_y + \phi_y \phi_y \phi_x + \phi_y \phi_x \phi_y + \phi_y \phi_x \phi_x) \bar{\mathbb{E}}_t \{ \bar{\mathbb{E}}_{t+1} [\bar{\mathbb{E}}_{t+2}(\theta_{t+3})] \} \\ &+ (\phi_x \phi_x + \phi_x \phi_y) \bar{\mathbb{E}}_t(\theta_{t+2}) + (\phi_y \phi_x + \phi_y \phi_y) \bar{\mathbb{E}}_t[\bar{\mathbb{E}}_{t+1}(\theta_{t+2})] + (\phi_x + \phi_y) \bar{\mathbb{E}}_t(\theta_{t+1}) + \theta_t. \end{aligned} \quad (\text{B.43})$$

Note that, together with the direct average expectation $\bar{\mathbb{E}}_t(\theta_{t+3})$, and the expectation of third order (the average expectation of the average expectation of the average expectation) $\bar{\mathbb{E}}_t \{ \bar{\mathbb{E}}_{t+1} [\bar{\mathbb{E}}_{t+2}(\theta_{t+3})] \}$, expression (B.43) also displays the average expectation at t of the average expectation at $t+2$, $\bar{\mathbb{E}}_t[\bar{\mathbb{E}}_{t+2}(\theta_{t+3})]$. Substituting further, other combinations of higher order expectations compounding appear too. In summary, equation (B.43) shows that, in presence of dispersed information, the pattern of higher order expectations can be extremely cumbersome, and so the requirement of the canonical approach to work out each possible combination of expectations quickly becomes prohibitive.

B.6 REAL BUSINESS CYCLE APPLICATION: FULL MODEL AND LOG-LINEARIZATION The economy is structured in a continuum of islands indexed by $i \in [0, 1]$. Each island is inhabited by a representative household i and by a representative firm i . Household i supplies labor services exclusively to firm i in a decentralized competitive labor

market. Labor of household i is the only labor productive in firm i . Households own capital in the economy and rent it out to firms in a centralized spot market. Capital, expressed in consumption goods, is productive in all the firms across the islands. The problem for Household i can be then written as

$$\max_{C_{it}, K_{it+i}^{(s)}} \mathbb{E}_{it} \left\{ \sum_{j=0}^{\infty} \beta^j \frac{C_{it+j}^{1-\frac{1}{\eta}} - 1}{1 - \frac{1}{\eta}} \right\}$$

with $\beta \in (0, 1)$, $\eta > 0$, and subject to a sequence of budget constraints of the form

$$C_{it} + K_{it+i}^{(s)} - (1 - \delta) K_{it}^{(s)} = W_{it} + R_t K_{it}^{(s)}, \quad t = 0, 1, 2, \dots,$$

where W_{it} is the wage rate in the labor market of island i and R_t is the rental rate of capital in the centralized capital market. Households are assumed, for the moment, to supply labor $N_{it}^{(s)}$ inelastically at the prevailing wage rate. We normalize the labor supplied by household i to $N_{it}^{(s)} = 1$. $K_{it+i}^{(s)}$ denotes the total capital that household i is bringing into period $t + 1$. The superscript (s) stands for “supply” to denote the fact that the capital that household i is bringing into period $t + 1$ will be the amount supplied by the same household in the centralized rental capital market at $t + 1$. Symmetrically, in what follows the superscript (d) will stand for demand.

The problem for the representative firm in island i is

$$\max_{N_{it}^{(d)}, K_{it}^{(d)}} Y_{it} - W_{it} N_{it}^{(d)} - R_t K_{it}^{(d)}$$

where

$$Y_{it}^{(s)} = Z_{it} \left(K_{it}^{(d)} \right)^{\alpha} \left(N_{it}^{(d)} \right)^{1-\alpha}, \quad \alpha \in (0, 1).$$

Output Y_{it} is supplied in the centralized market for output. In other words, in this economy there is only one consumption good centrally traded. The price of the consumption good at t is normalized to 1. Output is produced by firm i according to a Cobb-Douglas technology with labor and capital inputs and a technological factor Z_{it} that can be specified as

$$Z_{it} = e^{a_t + \varepsilon_{it}}.$$

The term a_t is common across all the islands, while ε_{it} is a productivity component that is specific to island i . The existence of a decentralized labor market together with an island specific productivity results in a labor income with an idiosyncratic risk component against which households would like to insure. We assume that markets for state contingent securities are not available, so that household i has to bear the labor income risk. We will also assume that the idiosyncratic labor income risk is not present in steady state, which means that the wealth distribution of the economy in steady state would be degenerate and an economy-wide representative household will exist. This is relevant at the linearization stage, as one would want to linearize the first order conditions for each island around the same steady state.

The first order conditions for firm i are

$$(1 - \alpha) Y_{it}^{(s)} = N_{it}^{(d)} W_{it}$$

and

$$\alpha Y_{it}^{(s)} = K_{it}^{(d)} R_t,$$

so that the wage bill in island i is equal to a fraction $(1 - \alpha)$ of output in the island, while the capital bill is the remaining fraction α . In addition, under the assumption that $N_{it}^{(s)} = 1$, market clearing for the decentralized labor market in island i , $N_{it}^{(s)} = N_{it}^{(d)}$, implies

$$Y_{it}^{(s)} = Z_{it} \left(K_{it}^{(d)} \right)^{\alpha} \quad \text{and} \quad (1 - \alpha) Y_{it}^{(s)} = W_{it}.$$

As a consequence, the participation of household i to the labor market, i.e. the observation of W_{it} , will result in the

knowledge of $Y_{it}^{(s)}$. This will always be the case, independently of the equilibrium behavior of the variables.

The Euler equation for household i is

$$1 = \mathbb{E}_{it} \left[\beta \left(\frac{C_{it}}{C_{it+1}} \right)^{\frac{1}{\eta}} (R_t + (1 - \delta)) \right],$$

while, using the first order conditions from firm i and the market clearing for the labor market, the budget constraint can be re-written as

$$C_{it} + K_{it+1}^{(s)} = (1 - \alpha) Y_{it}^{(s)} + (R_t + (1 - \delta)) K_{it}^{(s)}.$$

In addition, a no-Ponzi condition is assumed so that the solution path to the steady state has to satisfy the usual transversality condition.

To close the model in terms of market interactions we need to specify the market clearing condition for capital and for output, formally

$$\int_0^1 K_{it}^{(d)} di = \int_0^1 K_{it}^{(s)} di$$

and

$$\int_0^1 C_{it} di + \int_0^1 (K_{it+1}^{(s)} - (1 - \delta) K_{it}^{(s)}) di = \int_0^1 Y_{it}^{(s)} di.$$

We will work with a log-linearized version of the economy around a steady state that is derived under the assumption that the long run unconditional average of Z_{it} is 1. Notice that this implies that the economy does not display a growth trend in steady state. This is without loss of generality for the purpose of the application.

For any variable X_t we define x_t as $X_t = X^* e^{x_t}$, where X^* is the steady state value of X_t . The log-linearized economy is given by the following set of equations (details available from the authors upon request). The budget constraint for household i is

$$\left(\frac{1}{\alpha} (\rho + \delta) - \delta \right) c_{it} + k_{it+1}^{(s)} = \frac{1 - \alpha}{\alpha} (\rho + \delta) y_{it}^{(s)} + (\rho + \delta) r_t + \frac{1}{\beta} k_{it}^{(s)} \quad (\text{B.44})$$

where $\rho = \frac{1}{\beta} - 1$ is the rate of time preference. The Euler equation for household i is

$$\mathbb{E}_{it} [(c_{it} - c_{it+1}) + \eta\beta (\rho + \delta) r_{t+1}] = 0. \quad (\text{B.45})$$

Output supplied by firm i is given by

$$y_{it}^{(s)} = a_t + \varepsilon_{it} + \alpha k_{it}^{(d)}, \quad (\text{B.46})$$

and the rental rate for capital is

$$r_t = a_t + \varepsilon_{it} + (\alpha - 1) k_{it}^{(d)}. \quad (\text{B.47})$$

The demand for capital of firm i is

$$k_{it}^{(d)} = w_{it} - r_t^K \quad (\text{B.48})$$

and the wage rate is

$$w_{it} = y_{it}^{(s)}. \quad (\text{B.49})$$

The market clearing condition for aggregate capital is

$$\int_0^1 k_{it}^{(d)} di = \int_0^1 k_{it}^{(s)} di = k_t \quad (\text{B.50})$$

while the market clearing for aggregate output is

$$\left(\frac{1}{\alpha} (\rho + \delta) - \delta \right) c_t + (k_{t+1} - (1 - \delta) k_t) = \frac{1}{\alpha} (\rho + \delta) y_t \quad (\text{B.51})$$

where

$$c_t = \int_0^1 c_{it} di, \quad y_t = \int_0^1 y_{it}^{(s)} di.$$

The set of equations (B.44)-(B.51) completely describe the equilibrium dynamics of the linearized economy, conditional on the sequence of cross sectional distributions of information sets implicit in the expectational operator \mathbb{E}_{it} for $i \in [0, 1]$ and $\forall t$. Setting $\delta = 1$, one obtains the model equations of Section 5.

B.7 APPLICATIONS In this section we present examples of four applications that can be cast into our model specification (10)-(7). The list is by no means exhaustive.

EXAMPLE 1: REAL BUSINESS CYCLE WITH CAPITAL. This example corresponds to the application in the main text. In a standard real business cycle model with capital, in presence of dispersed information about the aggregate productivity shock and incomplete insurance markets, the linear dynamics of capital around the steady state can be expressed as

$$\alpha\beta\mathbb{E}_{it}(k_{it+2}) + \eta(1 - \alpha\beta)\mathbb{E}_{it}(r_{t+1}) - \mathbb{E}_{it}(a_{it+1}) = \alpha(1 + \beta)k_{it+1} - \alpha k_{it} - a_{it}, \quad (\text{B.52})$$

which is a standard second-order difference equation in capital, and where

$$r_{t+1} = \int_0^1 a_{it+1}\mu(i)di - (1 - \alpha) \int_0^1 k_{it+1}\mu(i)di, \quad (\text{B.53})$$

is the market-clearing rental rate for capital. Here β is the subjective discount factor, α is the capital share in the Cobb-Douglas output good technology, η is the elasticity of intertemporal substitution, and a_{it} is the individual productivity shock.¹⁶ Model (B.52)-(B.53) maps into (7)-(10) by setting $x_{it} = k_{it+1}$, $y_t = r_t$, $\theta_{it} = a_{it}$, and

$$\phi = \begin{pmatrix} \alpha\beta & \eta(1 - \alpha\beta) & -1 \end{pmatrix}, \psi(L) = \begin{pmatrix} \alpha(1 + \beta) - \alpha L & 0 & -1 \end{pmatrix}, \gamma(L) = \begin{pmatrix} (1 - \alpha)L & 1 & -1 \end{pmatrix}.$$

EXAMPLE 2: CALVO PRICING AND NEW KEYNESIAN PHILLIPS CURVE. Nimark (2008) considers a dispersed information version of the optimal Calvo pricing problem that micro-founds the New Keynesian Phillips Curve popularized Woodford (2003b) and Galí (2008). Nimark (2008) shows that the optimal price for an optimizing firm is

$$p_{it}^* = (1 - \beta\vartheta)(p_t + mc_{it}) + \beta\vartheta\mathbb{E}_{it}(p_{it+1}^*), \quad (\text{B.54})$$

where p_t is the aggregate price level, defined as $p_t = \vartheta p_t^* + (1 - \vartheta)p_{t-1}$, with $p_t^* \equiv \int_0^1 p_{it}^*\mu(i)di$, and mc_{it} is the individual marginal cost at time t specified as $mc_{it} = mc_t + v_{it}$ so that $\int_0^1 mc_{it}\mu(i)di = mc_t$. The parameter β is the discount factor for price setters, while ϑ measures the probability of resetting ones' price in a given period. Define $p_{it} \equiv \vartheta p_{it}^* + (1 - \vartheta)p_{t-1}$, which maintains $\int_0^1 p_{it}\mu(i)di = p_t$. The individual and aggregate price level dynamics can then be written as,

$$\beta\vartheta\mathbb{E}(p_{it+1}|\Omega_{it}) = p_{it} - \vartheta(1 - 2\vartheta\beta + \beta)p_t + (1 - \vartheta)p_{t-1} - \vartheta(1 - \beta\vartheta)mc_{it}. \quad (\text{B.55})$$

with

$$p_t = \int_0^1 p_{it}\mu(i)di. \quad (\text{B.56})$$

Equations (B.55)-(B.56) maps into (7)-(10) by setting $x_{it} = p_{it}$, $y_t = p_t$, $\theta_{it} = mc_{it}$, and

$$\phi = \begin{pmatrix} \beta\vartheta & 0 & 0 \end{pmatrix}, \psi(L) = \begin{pmatrix} 1 & -\vartheta(1 - 2\vartheta\beta + \beta) + (1 - \vartheta)L & -\vartheta(1 - \beta\vartheta) \end{pmatrix}, \gamma(L) = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}.$$

As recognized by Nimark (2008), in the presence of dispersed information a compact representation of the New Keynesian Phillips Curve cannot be obtained. However, once a solution for p_t is derived from (B.55)-(B.56), inflation dynamics are immediately given by $\pi_t = p_t - p_{t-1}$.

EXAMPLE 3: DYNAMIC ASSET PRICING. Singleton (1987) presents a dynamic asset pricing model motivated by the market microstructure of the U.S. bond market, which features a competitive, Walrasian market structure with a single security that is traded among speculative investors and nonspeculative or liquidity traders at the price

¹⁶Equations (B.52) and (B.53) are derived as part of the application of Section 5. See that section for details.

p_t .¹⁷ The security is assumed to pay a constant coupon every period, which we normalize to zero. Purchases of the security are financed by borrowing at the constant rate r , and the wealth of investor i evolves according to $w_{it+1} = z_{it}p_{t+1} - (1+r)(z_{it}p_t - w_{it})$. The i^{th} investor is assumed to have a one-period investment horizon and to rank alternative investment strategies according to the utility $\mathbb{E}_{it}[-\exp(-\varrho w_{it+1})]$, where ϱ is the constant coefficient of absolute risk aversion. Singleton (1987) shows that the demand schedule for the risky asset takes the form

$$z_{it} = \frac{1}{\varrho\nu} \mathbb{E}_{it}(p_{t+1}) - \frac{1+r}{\varrho\nu} p_t, \quad (\text{B.57})$$

where ν is the variance of p_{t+1} and is set to be an exogenous constant. Singleton (1987) assumes that the net supply of the asset, denoted by n_t , is specified as

$$n_t = f_t + \vartheta p_t. \quad (\text{B.58})$$

The shock to net asset supply f_t arises from nonspeculative traders (such as the U.S. Treasury, the Federal Reserve, financial intermediaries), that attempt to satisfy macroeconomic objectives for technical reasons related to the intermediation process. Nonspeculative traders are assumed to respond positively to an increase in prices; thus $\vartheta > 0$. Investors in setting their strategy z_{it} are assumed to receive a private signal, $f_{it} = f_t + v_{it}$, about the shock to the net asset supply. Market clearing is therefore given by

$$\int_0^1 z_{it}\mu(i)di = n_t. \quad (\text{B.59})$$

Model (B.57)-(B.59) maps into (7)-(10) by setting $x_{it} = z_{it}$, $y_t = p_t$, $\theta_{it} = f_{it}$, and

$$\phi = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \psi(L) = \begin{pmatrix} \varrho\nu & 1+r & 0 \end{pmatrix}, \gamma(L) = \begin{pmatrix} 1 & -\vartheta & -1 \end{pmatrix}.$$

EXAMPLE 4: CLASSICAL MONETARY MODELS OF INFLATION. In classical monetary models of inflation, money demand takes the form popularized by Cagan (1956),

$$m_{it} - p_t = -\alpha(\mathbb{E}_{it}(p_{t+1}) - p_t), \quad (\text{B.60})$$

where m_{it} is nominal money demand by agent i , p_t is an aggregate price index, and $\alpha > 0$. The money supply M_t is assumed to possess persistent dynamics specified as

$$M_t = \rho M_{t-1} + f_t, \quad (\text{B.61})$$

where f_t is a money supply shock process. The money market clearing condition is then

$$M_t = \int_0^1 m_{it}\mu(i)di \quad (\text{B.62})$$

Agents are assumed to receive a private signal, $f_{it} = f_t + v_{it}$, about the money supply shock. Equations (B.60)-(B.62) map into (7)-(10) by setting $x_{it} = m_{it}$, $y_t = p_t$, $\theta_{it} = f_{it}$, and

$$\phi = \begin{pmatrix} 0 & -\alpha & 0 \end{pmatrix}, \psi(L) = \begin{pmatrix} 1 & -(1+\alpha) & 0 \end{pmatrix}, \gamma(L) = \begin{pmatrix} 1 - \rho L & 0 & -1 \end{pmatrix}.$$

B.8 EXOGENOUS NOISE AND CONFOUNDING DYNAMICS In this section we show that confounding dynamics can be modeled also in the presence of exogenous signals that make the signal structure “rectangular,” as opposed to “square,” which is the case considered in the main analysis. There are two ways to preserve heterogeneous information in equilibrium—by continually adding exogenous noise until the noise terms overwhelm all signals, and/or by proving that there exists a zero inside the unit circle of the equilibrium as is done in Theorem 1. These categories are not

¹⁷Several papers have since used a similar setup to study a broad range of asset pricing issues (e.g., Bacchetta and van Wincoop (2006)).

mutually exclusive. Combinations of the two can certainly exist. In this section we first show that the standard way of introducing exogenous aggregate noise will *not* lead to the characteristic over- and under-reaction of the impulse response which is the hallmark of confounding dynamics.¹⁸ We then show that confounding dynamics can coexist with superimposed exogenous noise, and when they do, the characteristic over- and under-reaction reemerges. For transparency, we work within the stylized version of the generic rational expectations model [$\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, and $\phi_x = 0$, so that $\zeta = \phi_y$], and we modify the private signal of an arbitrary agent i to,¹⁹

$$\varepsilon_{it} = \varepsilon_t + v_{it}. \quad (\text{B.63})$$

All agents also observe the endogenous variable with superimposed exogenous noise $\tilde{\eta}_t$,

$$\tilde{y}_t = y_t + \tilde{\eta}_t. \quad (\text{B.64})$$

The noise $\tilde{\eta}_t$ is assumed to be of the form $\tilde{\eta}_t = U(L)\eta_t$, where $U(L)$ is a ratio of two lag polynomials in non-negative powers of L , and η_t is i.i.d. Gaussian with distribution $N(0, \sigma_\eta)$. Define the following relative signal-to-noise ratios,

$$\tau_\eta \equiv \frac{1}{1 + \sigma_\eta^2/\sigma_v^2 + \sigma_\eta^2/\sigma_\varepsilon^2}, \quad \tau_v \equiv \frac{1}{1 + \sigma_v^2/\sigma_\eta^2 + \sigma_v^2/\sigma_\varepsilon^2}, \quad (\text{B.65})$$

and note that $\lim_{\sigma_\eta \rightarrow \infty} \tau_v = \tau = \sigma_\varepsilon^2/(\sigma_\varepsilon^2 + \sigma_\eta^2)$. Following our solution strategy, we posit a candidate solution $y_t = Q_\varepsilon(L)\varepsilon_t + Q_\eta(L)\eta_t$. In order to achieve a closed-form solution, we follow Taub (1989) in specifying $U(L) = Q_\varepsilon(L) - Q_\eta(L)$, such that $\tilde{y}_t = Q_\varepsilon(L)(\varepsilon_t + \eta_t)$.²⁰ The following proposition characterizes analytically a rational expectations equilibrium for the exogenous noise economy, without confounding dynamics.

Proposition B2. *Consider model (7)-(10) with Assumption 1 and let $\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, and $\phi_x = 0$, so that $\zeta = \phi_y$. Let the information sets be specified as $\Omega_{it} = \varepsilon_t^i \vee \tilde{y}_t^i$. Define $\lambda(L) \equiv (LA(L) - \zeta\tau_v A(\zeta\tau_v))/(L - \zeta\tau_v)$, and let $U(L) = \lambda(L)$. There exists a unique rational expectations equilibrium,*

$$y_t = \left(\lambda(L) + \zeta\tau_\eta \frac{\lambda(L) - \lambda(\zeta)}{L - \zeta} \right) \varepsilon_t + \zeta\tau_\eta \left(\frac{\lambda(L) - \lambda(\zeta)}{L - \zeta} \right) \eta_t. \quad (\text{B.66})$$

which does not yield confounding dynamics, provided that,

$$\lambda(z) + \zeta\tau_\eta \frac{\lambda(z) - \lambda(\zeta)}{z - \zeta}, \quad (\text{B.67})$$

is invertible for all z inside the unit circle.

Proof. The first step in the proof is to recognize that the expectations of agent i in equilibrium take the form of a linear combination of current and past realizations of the observed variables ε_{it} and \tilde{p}_t ,

$$\mathbb{E}_{it}(y_{t+1}) = \hat{\pi}_1(L)\varepsilon_{it} + \hat{\pi}_2(L)\tilde{y}_t, \quad (\text{B.68})$$

where $\hat{\pi}_1(L)$ and $\hat{\pi}_2(L)$ are assumed to be representable as ratios of two finite-degree lag polynomials with zeros

¹⁸Superimposing exogenous noise is a common practice in most of the recent (and past) literature on dispersed information and aggregate fluctuations [e.g., Grossman and Stiglitz (1980), Wang (1993), Makarov and Rytchkov (2012), Angeletos and La'O (2013)].

¹⁹Assuming the private signal is ε_{it} , rather than θ_{it} , greatly simplifies the algebra in characterization of the equilibria of Propositions B2 and B3. All the key steps in the equilibrium derivation would go through if one were to consider θ_{it} .

²⁰The reasons for choosing a convenient form for $U(L)$ are two-fold. First, it streamlines the analytical derivation of the canonical factorization of the variance-covariance matrix. Second, it ensures that a solution to the equilibrium exists that takes the form of a finite order *ARMA* representation, or, in the frequency domain jargon, of an analytic function that can be represented as the ratio of two polynomials. The conditions for the existence of an ARMA solution in presence of exogenous noise superimposed to endogenous variables is an open active area of research, see Huo and Takayama (2016).

outside the unit circle. The guess for the equilibrium price is specified as

$$y_t = Q_\varepsilon(L)\varepsilon_t + Q_\eta(L)\eta_t. \quad (\text{B.69})$$

Substituting (B.68) in the equilibrium equation (7) and rearranging one obtains

$$y_t = \frac{\pi_1(L)}{1 - \pi_2(L)}\varepsilon_t + \frac{\pi_2(L)}{1 - \pi_2(L)}\tilde{\eta}_t, \quad (\text{B.70})$$

where $\pi_1(L) \equiv \beta\hat{\pi}_1(L) + A(L)$, and $\pi_2(L) \equiv \beta\hat{\pi}_2(L)$. Recalling that $\tilde{y}_t = y_t + \tilde{\eta}_t$, the information set of agent i can be then expressed as

$$\begin{pmatrix} \varepsilon_{it} \\ \tilde{y}_t \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ \frac{\pi_1(L)}{1 - \pi_2(L)} & 0 & \frac{U(L)}{1 - \pi_2(L)} \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ v_{it} \\ \eta_t \end{pmatrix} = \Xi(L) \begin{pmatrix} \varepsilon_t \\ v_{it} \\ \eta_t \end{pmatrix}. \quad (\text{B.71})$$

Denoting the entire history of the signal vector by ω , under the assumption that $U(L) = \pi_1(L)$ the variance-covariance generating function for the signal vector is

$$g_{ss}(z) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_v^2 & \frac{\pi_1(z^{-1})}{1 - \pi_2(z^{-1})}\sigma_\varepsilon^2 \\ \frac{\pi_1(z)}{1 - \pi_2(z)}\sigma_\varepsilon^2 & \frac{\pi_1(z)\pi_1(z^{-1})}{(1 - \pi_2(z))(1 - \pi_2(z^{-1}))}(\sigma_\varepsilon^2 + \sigma_\eta^2) \end{pmatrix}. \quad (\text{B.72})$$

In addition, the covariance generating function between the signal vector and y , the variable to be predicted, can be written as

$$g_{ys}(z) = \begin{pmatrix} \frac{\pi_1(z)}{1 - \pi_2(z)}\sigma_\varepsilon^2 & \frac{\pi_1(z)\pi_1(z^{-1})}{(1 - \pi_2(z))(1 - \pi_2(z^{-1}))}(\sigma_\varepsilon^2 + \pi_2(z)\sigma_\eta^2) \end{pmatrix}. \quad (\text{B.73})$$

Applying the usual Wiener-Kolmogorov prediction formula one sees that

$$\begin{bmatrix} \hat{\pi}_1(L) & \hat{\pi}_2(L) \end{bmatrix} = \left[L^{-1}g_{p\omega}(L)(\Xi^*(L^{-1})^T)^{-1} \right]_+ \Xi^*(L)^{-1}, \quad (\text{B.74})$$

where $\Xi^*(z)$ is the canonical factorization of the variance-covariance matrix $g_{ss}(z)$ such that $g_{ss}(z) = \Xi^*(z)\Xi^*(z^{-1})^T$. Following Rozanov (1967) and Taub (1989), the factorization can be shown to take the form

$$\Xi^*(z) = \begin{pmatrix} \sigma_z & \frac{\sigma_\varepsilon^2}{\sigma_y} \\ 0 & \frac{\pi_1(z)}{1 - \pi_2(z)}\sigma_y \end{pmatrix}, \quad (\text{B.75})$$

where $\sigma_y^2 \equiv \sigma_\varepsilon^2 + \sigma_\eta^2$, and $\sigma_z^2 \equiv \sigma_v^2 + \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_\varepsilon^2 + \sigma_\eta^2}$. Using (B.73) and (B.75) one can show that

$$\hat{\pi}_1(L) = \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_y^2\sigma_z^2} \left(\frac{\pi_1(L)}{L} - \frac{\pi_1(0)}{L} \right), \quad (\text{B.76})$$

and

$$\hat{\pi}_2(L) = \frac{1 - \pi_2(L)}{\pi_1(L)} \left[-\frac{\sigma_\eta^2}{\sigma_y^2} \left(\frac{\pi_1(L)}{L} - \frac{\pi_1(0)}{L} \right) - \frac{\pi_1(0)}{L(1 - \pi_2(0))} \right] + \frac{1}{L}. \quad (\text{B.77})$$

Given our definition of $\pi_1(L)$, equation (B.76) is a fixed point equation that can be solved independently for $\hat{\pi}_1(L)$. Define $\tau_v \equiv \frac{\sigma_\eta^2\sigma_\varepsilon^2}{\sigma_\eta^2\sigma_v^2 + \sigma_\varepsilon^2\sigma_\eta^2 + \sigma_\eta^2\sigma_\varepsilon^2} = \frac{\sigma_\varepsilon^2\sigma_\eta^2}{\sigma_y^2\sigma_z^2}$, the fixed point condition results in

$$\hat{\pi}_1(L) = \tau_v \frac{-\beta\hat{\pi}_1(0) + A(L) - A(0)}{L - \tau_v\beta}. \quad (\text{B.78})$$

Because $\tau_v\beta < 1$, to ensure covariance-stability of $\hat{\pi}_1(L)$ we need to pick $\hat{\pi}_1(0)$ so to cancel the unstable root at the denominator. This is achieved by setting $\hat{\pi}_1(0) = \beta^{-1}(A(\tau_v\beta) - A(0))$. Substituting this into (B.78), a closed form solution for $\hat{\pi}_1(L)$ is obtained. Using $\pi_1(L) = \beta\hat{\pi}_1(L) + A(L)$ one finally obtains

$$\pi_1(L) = \frac{LA(L) - \tau_v\beta A(\tau_v\beta)}{L - \tau_v\beta}. \quad (\text{B.79})$$

We let $\lambda(L) \equiv \pi_1(L)$ which agrees with the statement of Proposition B2. Turn now to condition (B.77). Using $\pi_2(L) = \beta\hat{\pi}_2(L)$ the fixed point condition can be expressed as

$$\frac{1}{1 - \beta\hat{\pi}_2(L)} = \frac{1}{\pi_1(L)} \left(\frac{L\pi_1(L) - \beta\kappa(L)}{L - \beta} \right), \quad (\text{B.80})$$

where

$$\kappa(L) \equiv \frac{\sigma_\eta^2}{\sigma_p^2} (\pi_1(L) - \pi_1(0)) + \frac{\pi_1(0)}{1 - \beta\hat{\pi}_2(0)}. \quad (\text{B.81})$$

Note that $\kappa(L)$ is a known function except for the constant $\hat{\pi}_2(0)$. Since $\beta < 1$, in order for the left hand side of (B.80) to be covariance-stationary, the right hand side should vanish at $L = \beta$. The constant $\hat{\pi}_2(0)$ can be conveniently chosen to achieve this by setting $\pi_1(\beta) - \kappa(\beta) = 0$. Solving this condition for $\hat{\pi}_2(0)$ and plugging the expression back into (B.80) one obtains

$$\frac{1}{1 - \beta\hat{\pi}_2(L)} = 1 + \beta\tau_\eta \frac{\pi_1(L) - \pi_1(\beta)}{\pi_1(L)(L - \beta)}, \quad (\text{B.82})$$

where $\tau_\eta \equiv \frac{\sigma_v^2 \sigma_\varepsilon^2}{\sigma_\eta^2 \sigma_v^2 + \sigma_v^2 \sigma_\varepsilon^2 + \sigma_\eta^2 \sigma_\varepsilon^2}$. Using $\pi_2(L) = \beta\hat{\pi}_2(L)$, one can use the resulting expression together with $\pi_1(L)$ to substitute in (B.69) and obtain (B.66) in Proposition B2. To complete the proof we need to argue why $\lambda(L) + \beta\tau_\eta \frac{\lambda(L) - \lambda(\beta)}{L - \beta} = 0$ must have no solution inside the unit circle. Note that, from (B.75), the determinant of the matrix $\Xi^*(L)$ is proportional to $\frac{\pi_1(L)}{1 - \pi_2(L)}$, and, for the matrix to be a canonical factorization, the determinant must not vanish inside the unit circle. Because $\frac{\pi_1(L)}{1 - \pi_2(L)} = \lambda(L) + \beta\tau_\eta \frac{\lambda(L) - \lambda(\beta)}{L - \beta}$, for $\Xi^*(L)$ to be the appropriate factorization, the right hand side must not vanish inside the unit circle. \square

The form of the equilibrium (B.66) can be best understood by studying the limiting functions of the noise terms. Note that the polynomial $\lambda(L)$ takes the form of a Hansen-Sargent formula involving $A(L)$ and $\tau_v \zeta$. To understand its role, suppose that the public information \check{y}_t is made uninformative so that $\tau_\eta \rightarrow 0$ (i.e. $\sigma_\eta \rightarrow \infty$). The equilibrium would then just be equal to $y_t = \lambda(L)\varepsilon_t$, which is the first term in (B.66) with τ_v equal to τ . As soon as public information is made informative two additional terms appear, one which captures the additional information about ε_t transmitted by the public information, and the other that injects the public noise η_t into the equilibrium price. Note that the two terms enter the equilibrium price with the same dynamics, which is a consequence of the assumption $U(L) = \lambda(L)$. This process is also characterized by a Hansen-Sargent formula involving $\lambda(L)$ and $\tau_\eta \zeta$. When public information is made arbitrarily precise, i.e. $\sigma_\eta \rightarrow 0$ so that $\tau_\eta = 1$ then (B.66) corresponds to the full information equilibrium.

A comparison with Theorem 1 reveals that the additional noise of Proposition B2 coming from (B.64) implies that condition (20), which guarantees heterogeneous beliefs are preserved in equilibrium, is no longer necessary. In fact, (B.67) is an explicit assumption that there are no zeros inside the unit circle, which is the standard assumption in models with exogenous noise. In turn, this implies that the equilibrium cannot support confounding dynamics.

To see this more clearly, suppose that $A(L) = 1 + \theta L$. Applying the results of Appendix A.1, the full information solution can be immediately obtained as the $MA(1)$ process,

$$y_t = (1 + \theta\zeta)\varepsilon_t + \theta\varepsilon_{t-1}. \quad (\text{B.83})$$

Substituting into the equilibrium (B.66) under the assumption that, $\theta < 1/(1 - \zeta(\tau_v + \tau_\eta))$, so that the invertibility of (B.67) holds, yields,

$$y_t = (1 + \theta\zeta(\tau_v + \tau_\eta))\varepsilon_t + \theta\varepsilon_{t-1} + \zeta\tau_\eta(1 + \theta)\eta_t. \quad (\text{B.84})$$

The impulse response dynamics of y_t in (B.84) to a shock ε_t are entirely consistent with the optimal prediction formula associated with the standard signal extraction problem described in Section 2. The impulse response to a shock in ε_t is smaller than the full information counterpart at impact, since $\tau_v + \tau_\eta < 1$, but otherwise unchanged (i.e. it matches the dashed dynamics of Figure 1).

Next we want to characterize a solution with both exogenous noise and confounding dynamics. Under the same assumptions about the private and public information signals, we posit a candidate solution $y_t = \tilde{Q}_\varepsilon(L)\varepsilon_t + \tilde{Q}_\eta(L)\eta_t$, and let $U(L) = \tilde{Q}_\varepsilon(L) - \tilde{Q}_\eta(L)$. The following proposition holds.

Proposition B3. *Consider model (7)-(10) and let $\tilde{\gamma}_y(L) = 1$, $\psi_x(L) = 1$, $\psi_y(L) = 0$, $\phi_x = 0$, so that $\zeta = \phi_y$. Let the information sets be specified as $\Omega_{it} = \varepsilon_i^t \vee \tilde{y}^t$. Define $\tilde{\lambda}(L) \equiv (1 - \lambda L)(LA(L) - \lambda A(\lambda))/(L - \lambda)$, for some real constant λ . There exists a Rational Expectations Equilibrium with Confounding Dynamics of the form*

$$y_t = \frac{L - \lambda}{1 - \lambda L} \left(\frac{(L - \zeta(1 - \tau_\eta))\tilde{\lambda}(L) - \tau_\eta\tilde{\lambda}(\zeta)}{(L - \zeta)(L - \zeta(1 - \tau_\eta)\tau)} \varepsilon_t + \tau_\eta \frac{(1 - \tau)\zeta\tilde{\lambda}(L) - (L - \zeta\tau_\eta)\tilde{\lambda}(\zeta)}{(1 - \tau)(L - \zeta)(L - \zeta(1 - \tau_\eta)\tau)} \eta_t \right), \quad (\text{B.85})$$

if, and only if, there exists a $\lambda \in (-1, 1)$ that solves

$$(1 - \tau)\tilde{\lambda}((1 - \tau_\eta)\tau\zeta) + \tau\tau_\eta\tilde{\lambda}(\zeta) = 0. \quad (\text{B.86})$$

Proof. The candidate solution for the equilibrium in Proposition B3 is specified as

$$y_t = \tilde{Q}_\varepsilon(L)\varepsilon_t + \tilde{Q}_\eta(L)\eta_t. \quad (\text{B.87})$$

The first part of the proof is equivalent to that of Proposition B2, up until equation (B.73). We then need to conjecture confounding dynamics, which we do by assuming that there exists a $\lambda \in (-1, 0)$ such that

$$\frac{\pi_1(\lambda)}{1 - \pi_2(\lambda)} = 0. \quad (\text{B.88})$$

If this is the case then the matrix $\Xi^*(z)$ in (B.75) harbors confounding dynamics since its determinant vanishes at λ . For notational convenience we assume that $\frac{\pi_1(\lambda)}{1 - \pi_2(\lambda)} = \pi(L)(L - \lambda)$, which embeds conjecture (B.88). To obtain the canonical factorization of $g_{ss}(z)$ we apply the steps in Appendix C.3 to $\Xi^*(z)$ and we obtain

$$\Xi^{**}(z) = \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_v^2}} \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_v^2 & 0 \\ \sigma_\varepsilon^2(z - \lambda)\pi(z) & \sigma_y\sigma_z\pi(z)(1 - \lambda z) \end{pmatrix}. \quad (\text{B.89})$$

For convenience define $\tilde{\pi}_2(z) = (z - \lambda)\pi_2(z)$, and using $\Xi^{**}(z)$ in the Wiener-Kolmogorov formula (B.74) one obtains the following two fixed point conditions in $\pi(L)$ and $\tilde{\pi}_2(L)$ after some straightforward rearrangements:

$$\pi(L)[(L - \lambda)(L - \zeta\tau) - (1 - \tau)\tilde{\pi}_2(L)L] = \zeta\tau\lambda\pi(0) + A(L)L, \quad (\text{B.90})$$

and

$$\frac{\tilde{\pi}_2(L)}{L - \lambda} = \frac{\zeta\tau_\eta(1 - \lambda L)h_1(L) - (L - \zeta\tau)\frac{\sigma_v^2}{\sigma_z^2\sigma_y^2}\frac{\zeta}{(1 - \tau)}h_2(L)}{(L - \zeta(1 - \tau_\eta))(1 - \lambda L)h_1(L) - \zeta\frac{\sigma_v^2}{\sigma_z^2\sigma_y^2}h_2(L)}, \quad (\text{B.91})$$

where

$$h_1(L) \equiv \zeta\tau\lambda\pi(0) + A(L)L, \quad (\text{B.92})$$

and

$$h_2(L) \equiv \pi(0)(L - \lambda) \left(\tau_\eta - \tilde{\pi}(0)\frac{\sigma_\eta^2}{\lambda} \right) - \frac{\sigma_\eta^2}{(1 - \tau)} \frac{1 - \lambda^2}{\lambda} \left(A(\lambda) + \zeta\tau\pi(0) \right). \quad (\text{B.93})$$

Substituting (B.91) into (B.90) one obtains

$$\pi(L)(L - \lambda) = \frac{(L - \zeta(1 - \tau_\eta))(1 - \lambda L)h_1(L) - \zeta\frac{\sigma_v^2}{\sigma_z^2\sigma_y^2}Lh_2(L)}{(1 - \lambda L)(L - \zeta)(L - \zeta\tau(1 - \tau_\eta))}. \quad (\text{B.94})$$

We require $\pi(L)(L - \lambda)$ to be stationary, which means that the two unstable roots at the denominator, $\zeta < 1$ and $\zeta\tau(1 - \tau_\eta) < 1$, need to be removed. In addition, our conjecture of confounding dynamics requires the left hand side expression to vanish at $L = \lambda$. We can achieve all this by the appropriate choice of constants $\pi(0)$, $\tilde{\lambda}_2(0)$ and λ . We thus have the following three conditions in three unknowns,

$$(\lambda - \zeta(1 - \tau_\eta)(1 - \lambda^2)h_1(\lambda) - \zeta \frac{\sigma_v^2}{\sigma_z^2 \sigma_y^2} \lambda h_2(\lambda) = 0, \quad (\text{B.95})$$

$$\sigma_\varepsilon^2(1 - \lambda\zeta)h_1(\zeta) - \zeta h_2(\zeta) = 0, \quad (\text{B.96})$$

$$(1 - \zeta\tau(1 - \tau_\eta)\lambda)h_1(\zeta\tau(1 - \tau_\eta)) + \zeta \frac{\sigma_\varepsilon^2}{\sigma_z^2 \sigma_y^2} = 0. \quad (\text{B.97})$$

We first note that $\lambda h_2(\lambda) = \tau_\eta h_1(\lambda)(1 - \lambda^2)$, which implies that condition (B.95) is satisfied when $\pi(0) = \frac{A(\lambda)}{\zeta\tau}$. Substituting this into the expressions for $h_1(L)$ and $h_2(L)$ one sees that,

$$h_1(L) = LA(L) - \lambda A(\lambda), \quad \text{and} \quad h_2(L) = \frac{A(\lambda)}{\zeta\tau}(L - \lambda) \left(\tau_\eta - \tilde{\pi}(0) \frac{\sigma_\eta^2}{\lambda} \right). \quad (\text{B.98})$$

Using these expressions into (B.96) one obtains

$$\left(\tau_\eta - \tilde{\pi}(0) \frac{\sigma_\eta^2}{\lambda} \right) = \tau \sigma_\varepsilon^2 \frac{(1 - \lambda\zeta)}{A(\lambda)(\zeta - \lambda)} (\lambda A(\lambda) - \zeta A(\zeta)) \quad (\text{B.99})$$

Next define

$$\tilde{\lambda}(L) \equiv (1 - \lambda L) \frac{LA(L) - \lambda A(\lambda)}{L - \lambda}, \quad (\text{B.100})$$

and note that condition (B.97) is satisfied when

$$\tilde{\lambda}(\zeta\tau(1 - \tau_\eta)) + \tilde{\lambda}(\zeta) \frac{\sigma_\varepsilon^4}{\sigma_z^2 \sigma_y^2} = 0, \quad (\text{B.101})$$

which corresponds to (B.86) once we multiply both sides by $\sigma_v^2/\sigma_\varepsilon^2$. With some additional straightforward algebra is then possible to solve for $\pi(L)$ and $\pi_2(L)$, and using the conditions, $\tilde{Q}_\varepsilon(L) = \pi(L)$, and $\tilde{Q}_\eta(L) = \pi_1(L) \frac{\pi_2(L)}{1 - \pi_2(L)}$, equation (B.85) obtains. \square

The building block of this equilibrium is $\tilde{\lambda}(L)$, which can be directly compared to the $\lambda(L)$ function of Proposition B2. As discussed above, the second term of a Hansen-Sargent prediction formula has an informational interpretation in that it amounts to what must be subtracted away from a complete-information equilibrium. The conditioning down associated with $\tilde{\lambda}(L)$ of Proposition B3 is due to the endogenous zero, λ determined by (B.86); while the conditioning down associated with $\lambda(L)$ of Proposition B2 is due to the exogenous noise term, τ_v .

Most importantly, Proposition B3 reintroduces confounding dynamics. Under the specification $A(L) = 1 + \theta L$, the equilibrium (B.85), when (B.86) is satisfied,²¹ is given by

$$y_t = \frac{L - \lambda}{1 - \lambda L} \left((\theta - \lambda(1 + \lambda\theta) - \lambda\theta\zeta(\tau_\eta + (1 - \tau_\eta)\tau))\varepsilon_t - \lambda\theta\varepsilon_{t-1} - \zeta\tau_\eta\lambda\theta\eta_t \right), \quad (\text{B.102})$$

Comparing expression (B.102) to (B.84), both contain an $MA(1)$ term for ε_t , and a constant coefficient on η_t .

²¹Equilibria (B.66) and (B.85) do not necessarily exist simultaneously for the same parameter values. However, our objective here is to compare the qualitative features of the equilibrium dynamics across the two different classes of equilibria they represent, one with exogenous noise only, the other with both exogenous noise and confounding dynamics. For such exercise, the space of existence across parameter values has a secondary relevance.

However, for equilibrium (B.102), the $MA(1)$ term is multiplied by the factor $(L - \lambda)/(1 - \lambda L)$ which, as seen in Section 2, injects the dynamic pattern typical of confounding dynamics.²² The impulse response dynamics of y_t in (B.102) to a shock ε_t is smaller than the full information counterpart at impact and it matches the qualitative behavior of confounding dynamics in Figure 1.²³ Taken together, Propositions B2 and B3 show that it is the learning mechanism due to confounding dynamics, rather than the one due to exogenous noise, that injects persistence in innovations, and, simultaneously, an amplification pattern that resembles waves of optimism and pessimism.

²²For $\theta > 0$, one can show that $\lambda \in (-1, 0)$ when (B.86) is satisfied.

²³To see this one need to show that the impact coefficient in (B.102) is smaller than the impact coefficient in (B.83), which corresponds to $-\lambda(\theta - \lambda(1 + \lambda\theta) - \lambda\theta\zeta(\tau_\eta + (1 - \tau_\eta)\tau)) < 1 + \theta\zeta$. This can be easily shown using the property that when $\theta > 0$, $-\lambda\theta \in (0, 1)$.

C TECHNICAL PRELIMINARIES FOR FREQUENCY DOMAIN METHODS

Elementary results concerning the theory of stationary stochastic processes and the residue calculus are necessary for grasping the z -transform approach (also known as frequency domain approach) advocated in the main text. The purpose of this appendix is to offer readers unfamiliar with the methods used to derive our results the minimal background necessary so that the paper is self-contained. This appendix introduces important theorems that are relatively well known but is by no means exhaustive. Interested readers are directed to Brown and Churchill (2013) and Whittle (1983) for good references on complex analysis and stochastic processes. Sargent (1987) provides a good introduction to these concepts and discusses economic applications.

C.1 VARIANCE-COVARIANCE GENERATING FUNCTION Consider two-covariance stationary linear-Gaussian multivariate processes, $\{\omega_t, t \in \mathbb{Z}\}$ and $\{s_t, t \in \mathbb{Z}\}$, where the vector dimensions are $n \times 1$, and $m \times 1$, respectively. Let $\Upsilon_{\omega s}(j)$ denote the $m \times n$ unconditional covariance matrix between ω_t and s_{t-j} , for $j \in \mathbb{Z}$, formally

$$\Upsilon_{\omega s}(j) \equiv \mathbb{E}(\omega_t s_{t-j}^T) - \mathbb{E}(\omega_t) \mathbb{E}(s_{t-j}^T), \quad (\text{C.1})$$

where T denotes transpose. The variance-covariance generating function is then defined as

$$g_{\omega s}(z) \equiv \sum_{j=-\infty}^{\infty} \Upsilon_{\omega s}(j) z^j, \quad (\text{C.2})$$

where $g_{\omega s}(z)$ is an $m \times n$ matrix. When $\omega_t = s_t$ the function is referred to as the auto-covariance generating function and denoted by $g_{\omega\omega}(z)$, or $g_{ss}(z)$. An extensive treatment of the properties of the variance-covariance generating function can be found in Sargent (1987).

C.2 WOLD FUNDAMENTAL REPRESENTATION THEOREM Much of the analysis in the main text is conducted in the space of lag polynomials without specific functional forms assumed (e.g., $ARMA(m, n)$). The Wold representation theorem allows for such a general specification.

Theorem C1. [*Wold Representation Theorem*] Let $\{s_t\}$ be any $(n \times 1)$ covariance stationary stochastic process with $\mathbb{E}(s_t) = 0$. Then it can be uniquely represented in the form

$$s_t = \eta_t + \Gamma^*(L) \hat{w}_t \quad (\text{C.3})$$

where $\Gamma^*(L)$ is a matrix polynomial in the lag operator, and $\sum_{j=0}^{\infty} \Gamma_j^* \Gamma_j^{*T}$ is convergent. The process \hat{w}_t is n -variate white noise with $\mathbb{E}(\hat{w}_t) = 0$, $\mathbb{E}(\hat{w}_t \hat{w}_t') = I_n$ and $\mathbb{E}(\hat{w}_t \hat{w}_{t-m}') = 0$ for $m \neq 0$. The process $\Gamma_0^* \hat{w}_t$ is the innovation in predicting s_t linearly from its own past:

$$\Gamma_0^* \hat{w}_t = s_t - \mathcal{P}(s_t | s_{t-1}, s_{t-2}, \dots), \quad (\text{C.4})$$

where $\mathcal{P}(\cdot)$ denotes linear projection. The process η_t is linearly deterministic; there exists an $n \times 1$ vector c_0 and $n \times n$ matrices C_s such that without error $\eta_t = c_0 + \sum_{s=1}^{\infty} C_s \eta_{t-s}$ and $\mathbb{E}[\hat{w}_t \eta_{t-m}'] = 0$ for all m .

The Wold representation theorem states that *any* covariance stationary process can be written as a linear combination of a (possibly infinite) moving average representation where the innovations are the linear forecast errors for s_t and a process that can be predicted arbitrarily well by a linear function of past values of s_t . The theorem is a *representation* determined by second moments of the stochastic process only and therefore may not fully capture the data generating process. For example, that the decomposition is linear suggests that a process could be deterministic in the strict sense and yet linearly non-deterministic; Whittle (1983) provides examples of such processes. The innovations in the Wold representation are generated by linear projections which need not be the same as the conditional expectation ($\mathbb{E}[s_t | s_{t-1}, s_{t-2}, \dots]$). However, when working with linear Gaussian stochastic processes, as is standard in the rational expectations literature, the best conditional expectation coincides with linear projection.

As discussed in detail in the paper, the innovations derived from the Wold representation are an essential element of a rational expectations equilibrium in that they define the information set obtained from conditioning on current and past s_t 's. Using the language of Rozanov (1967), the innovation sequence $\{\hat{w}_{t-j}\}_{j=0}^{\infty}$ of (C.4) are fundamental for the sequence $\{s_{t-j}\}_{j=0}^{\infty}$ if the Hilbert space generated by the observables is equivalent (in mean-square norm) to the Hilbert space generated by the innovations.

C.3 CANONICAL FACTORIZATION Let s_t be specified as

$$s_t = \Gamma(L)u_t, \quad (\text{C.5})$$

where $\Gamma(L)$ is a lag polynomial matrix with square-summable coefficients, and u_t is a $m \times 1$ vector of i.i.d. Gaussian innovations with zero mean and variance-covariance matrix Σ_u . Testing for fundamentalness in a process like (C.5) can be done by checking for the invertibility of $\Gamma(z)$ for $|z| < 1$.²⁴ While testing for fundamentalness is straightforward, deriving the unique fundamental Wold representation is not. Here we rely on powerful factorization theorems. The following definition is taken from Whittle (1983). If $g(z)$ is a square matrix function of z , then the canonical factorization

$$g(z) = \Gamma^*(z)\Gamma^*(z^{-1})^T, \quad (\text{C.6})$$

will refer to a factorization in which both $\Gamma^*(z)$ and its inverse have valid expansions in non-negative powers of z for $|z| \leq 1$; and $\Gamma^*(z^{-1})^T$ and its inverse have expansions in non-positive powers of z for $|z| \geq 1$. The canonical factorization of the variance-covariance generating function delivers the Wold fundamental representation of the time series. It yields both the fundamental moving average Wold representation, and the autoregressive representation that is consistent with $\text{span}\{s_{t-j}\}_{j=0}^{\infty}$ being equivalent to $\text{span}\{\hat{w}_{t-j}\}_{j=0}^{\infty}$.

There does not exist a general method to compute analytically the canonical factorization of any arbitrary matrix $g(z)$. Several methods that have been proposed work well when the knowledge of the structure of the matrix $g(z)$ is used. Rozanov (1967), page 47, proposes an algorithm to obtain $\Gamma^*(z)$ from $\Gamma(L)$ that can be used when $\Gamma(L)$ is known to have a finite number of isolated singularities $\lambda_1, \lambda_2, \dots, \lambda_p$ inside the unit circle. In the case of $n = 2$, and one singularity at $|\lambda| < 1$, one has that

$$\Gamma^*(z) = \Gamma(z)W_\lambda B_\lambda(z), \quad (\text{C.7})$$

where $B_\lambda(z)$ is the Blaschke matrix

$$B_\lambda(z) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1-z\lambda}{z-\lambda} \end{bmatrix}. \quad (\text{C.8})$$

The constant matrix W_λ is a unitary matrix, so that $W_\lambda^{-1} = W_\lambda^T$, whose columns are the left singular vectors of the system matrix $\Gamma(z)$ evaluated at λ . While the inverse of $\Gamma(\lambda)$ does not exist, the pseudo-inverse (also known as the Moore-Penrose inverse) always exists and it is given by

$$\Gamma(\lambda)^+ = U\Sigma V^T, \quad (\text{C.9})$$

where $V\Sigma U^T$ is the singular value decomposition of the matrix $\Gamma(\lambda)$. The left singular vectors of $\Gamma(\lambda)$ are the columns of U . The following result, which can be showed by using widely available formulas for the singular value decomposition of 2 by 2 matrices is used in the proof of Lemma A1.

Lemma C1. *Consider the matrix*

$$N = \begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix}. \quad (\text{C.10})$$

Then the matrix U in the singular value decomposition of $N = V\Sigma U^T$, is

$$U = \frac{1}{\sqrt{c^2 + d^2}} \begin{bmatrix} \frac{cd}{|d|} & -\frac{cd}{|c|} \\ \frac{d^2}{|d|} & \frac{c^2}{|c|} \end{bmatrix}. \quad (\text{C.11})$$

²⁴There are many ways to test for fundamentalness, see Fernandez-Villaverde, Rubio, Sargent, and Watson (2007).

C.4 RIESZ-FISCHER THEOREM

Theorem C2. [*Riesz-Fischer*] Let $D(\sqrt{r})$ denote a disk in the complex plane of radius \sqrt{r} centered at the origin. There is an equivalence (i.e. an isometric isomorphism) between the space of r -summable sequences $\sum_j r^j |f_j|^2 < \infty$ and the Hardy space of analytic functions $f(z)$ in $D(\sqrt{r})$ satisfying the restriction

$$\frac{1}{2\pi i} \oint f(z) f(rz^{-1}) \frac{dz}{z} < \infty$$

where \oint denotes (counterclockwise) contour integration around $D(\sqrt{r})$. An analytic function satisfying the above condition is said to be r -integrable.²⁵

The Riesz-Fischer theorem ensures that one can work either in the space of infinite sequence of square-summable matrix coefficients $\{\dots, Q_{-1}, Q_0, Q_1, \dots\}$, or in the space of complex-valued analytic functions $Q(z) = \sum_{j=-\infty}^{\infty} Q_j z^j$, since the two spaces are equivalent.

C.5 WIENER-KOLMOGOROV PREDICTION FORMULA Consider the problem of computing the linear least-squares estimate for ω_{t+j} , denoted by $\hat{\omega}_{t+j}$, conditional on the realized history $\{s_{t+j}\}_{j=0}^{\infty}$. The solution to the problem consists in the sequence of real valued matrices $\{\Pi_j\}_{j=0}^{\infty}$, or, equivalently, the complex-valued function $\Pi(z)$, such that

$$\hat{\omega}_{t+j} = \sum_{j=0}^{\infty} \Pi_j s_{t-j}. \quad (\text{C.12})$$

We assume that all the processes have zero unconditional mean. The following result is a version of the Wiener-Kolmogorov prediction formula taken from Whittle (1983).

Theorem C3. Suppose that $g_{ss}(z)$ has the canonical factorization

$$g_{ss}(z) = \Gamma^*(z) \Gamma^*(z^{-1})^T. \quad (\text{C.13})$$

then the generating function $\Pi_j(z) = \sum_{i=0}^{\infty} \Pi_i z^i$ of the optimal estimate $\hat{\omega}_{t+j}$ is

$$\Pi_j(z) = [z^{-j} g_{\omega s}(z) (\Gamma^*(z^{-1})^T)^{-1}]_+ \Gamma^*(z)^{-1}. \quad (\text{C.14})$$

Proof. The minimization of the the means squared forecast errors $\mathbb{E}(\hat{\omega}_t - \omega_t)^2$ leads to the set of first order conditions

$$\sum_{k=0}^{\infty} \Upsilon_{ss}(i-k) = \Upsilon_{\omega s}(i-j), \quad \text{for } i = 0, 1, 2, \dots \quad (\text{C.15})$$

Multiplying both sides by z^i and adding over all integral i 's one gets the Wiener-Hopf relationship

$$\Pi_j(z) g_{ss}(z) = z^{-j} g_{\omega s}(z) + h(z), \quad (\text{C.16})$$

where $h(z) = \sum_{i=-\infty}^{-1} h_i z^i$, is an unknown matrix series in negative powers of z . If we post-multiply both sides by $[\Gamma^*(z^{-1})^T]^{-1}$, we obtain

$$\Pi_j(z) \Gamma^*(z) = z^{-j} g_{\omega s}(z) [\Gamma^*(z^{-1})^T]^{-1} + h(z) [\Gamma^*(z^{-1})^T]^{-1}, \quad (\text{C.17})$$

Note that both sides can be represented as matrix series in powers of z whose coefficients have to obey the above relationship. By construction, the left hand side has only positive powers of z , while the second term of the right hand side has only negative powers of z . If we apply the annihilating operator on both sides we thus get

$$\Pi_j(z) \Gamma^*(z) = [z^{-j} g_{\omega s}(z) [\Gamma^*(z^{-1})^T]^{-1}]_+. \quad (\text{C.18})$$

²⁵This theorem is usually proved for the case $r = 1$ and for functions defined on the boundary of a disk. For further exposition see Sargent (1987).

The last step consists in post-multiplying both sides by $\Gamma^*(z)^{-1}$ so that (C.14) obtains. \square

The requirement that $\Gamma^*(z)$ should be the canonical factorization of $g_{ss}(z)$ is essential in two steps. First, from it we ensure that $[\Gamma^*(z^{-1})^T]^{-1}$ has an expansion in non-positive powers of z , which means that the term $h(z)[\Gamma^*(z^{-1})^T]^{-1}$ disappears when the annihilating operator is applied. Second, we also ensure that $\Gamma^*(z)^{-1}$ has an expansion in non-negative powers of z , which result in $\Pi(z)$ having an expansion in non-negative powers of z only, as required.

C.6 WHITEMAN (1983) SOLUTION METHOD: EXISTENCE AND UNIQUENESS OF REE We use the existence and uniqueness criteria of Whiteman (1983) developed for linear, rational expectations equilibria. The following works through the three relevant cases—existence but no uniqueness, no existence, and existence-uniqueness. Consider the following generic rational expectations model

$$\zeta\tilde{\zeta}\mathbb{E}_t y_{t+1} - (\zeta + \tilde{\zeta})y_t + y_{t-1} = \theta_t, \quad \theta_t = A(L)\varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} N(0, 1) \quad (\text{C.19})$$

where ε_t is assumed to be fundamental for θ_t (i.e., $A(L)$ is assumed to have a one-sided inverse in non-negative powers of L). Following the solution principle, we will look for a solution that is square-summable in the Hilbert space generated by the fundamental shock ε , $y_t = Q(L)\varepsilon_t$ (third tenet). If we invoke the optimal prediction formula (C.14), then $\mathbb{E}_t y_{t+1} = [Q(L)/L]_+ \varepsilon_t = L^{-1}[Q(L) - Q_0]\varepsilon_t$. Together with the fourth tenet of the solution principle (i.e., that the rational expectation restrictions hold for all realizations of ε), this implies that (C.19) can be written in z -transform as

$$z^{-1}[Q(z) - Q_0]\zeta\tilde{\zeta} - (\zeta + \tilde{\zeta})Q(z) + zQ(z) = A(z)$$

Multiplying by z and rearranging delivers

$$Q(z) = \frac{zA(z) + Q_0}{(\zeta - z)(\tilde{\zeta} - z)} \quad (\text{C.20})$$

Appealing to the Riesz-Fischer Theorem, square-summability (stationarity) in the time domain is tantamount to analyticity of $Q(z)$ on the unit disk. The function $Q(z)$ is analytic at z_0 if it is continuously (complex) differentiable in an open neighborhood of z_0 .²⁶ Any rational function ($f(z)/g(z)$) where $f(\cdot)$ and $g(\cdot)$ are polynomials will be analytic on the unit disk provided $g(z) \neq 0$ at any point inside the unit circle. The extent to which this is true for $Q(z)$ depends upon the parameters ζ and $\tilde{\zeta}$.

As shown in Whiteman (1983), there are three cases one must consider. First, assume that $|\zeta| > 1$ and $|\tilde{\zeta}| > 1$. Then (C.20) is an analytic function on $|z| < 1$ and the representation is given by

$$y_t = \left[\frac{LA(L) + Q_0}{(\zeta - L)(\tilde{\zeta} - L)} \right] \varepsilon_t \quad (\text{C.21})$$

For any finite value of Q_0 , this is a solution that lies in the Hilbert space generated by $\{\theta_t\}$ and satisfies the tenets of the solution principle. Thus, we have existence but not uniqueness because Q_0 can be set arbitrarily.

The second case to consider is $|\tilde{\zeta}| < 1 < |\zeta|$. In this case, the function $Q(z)$ has an isolated singularity at ζ , implying that $Q(z)$ is not analytic on the unit disk. In this case, the free parameter Q_0 can be set to remove the singularity at ζ by setting Q_0 in such a way as to cause the residue of $Q(\cdot)$ to be zero at ζ

$$\lim_{z \rightarrow \zeta} (\tilde{\zeta} - z)Q(z) = \frac{\zeta A(\zeta) + Q_0}{\zeta - \tilde{\zeta}} = 0$$

Solving for Q_0 delivers $Q_0 = -\zeta A(\zeta)$. Substituting this into (C.21) yields the following rational expectations equi-

²⁶Analytic is synonymous with holomorphic, regular and regular analytic.

librium

$$y_t = \left[\frac{LA(L) - \zeta A(\zeta)}{(\zeta - L)(\tilde{\zeta} - L)} \right] \varepsilon_t. \quad (\text{C.22})$$

The function $Q(z)$ is now analytic and (C.22) is the unique solution that lies in the Hilbert space generated by $\{\theta_t\}$. The solution is the ubiquitous Hansen-Sargent prediction formula that clearly captures the cross-equation restrictions that are the “hallmark of rational expectations models” [Hansen and Sargent (1980)].²⁷

The final case to consider is $|\zeta| < 1$ and $|\tilde{\zeta}| < 1$. In this case, (C.20) has two isolated singularities at ζ and $\tilde{\zeta}$, and Q_0 cannot be set to remove both singularities.²⁸ Hence in this case, there is no solution in the Hilbert space generated by $\{\theta_t\}$ and we do not have existence.

C.7 ANNIHILATING OPERATOR Let $H(z) = \sum_{m=-\infty}^{+\infty} H_m z^m$. The following Lemma is due to Hansen and Sargent (1980).

Lemma C2. *Let $H(z)$ be a regular function in $|z| < 1$, with at most a finite number of singularities z_1, z_2, \dots, z_k in $|z| < 1$. Let $\pi_1(z), \pi_2(z), \dots, \pi_k(z)$ denote the principal parts of the Laurent series expansion of $H(z)$ around the singularities. Then*

$$[H(z)]_+ = H(z) - \sum_{j=1}^k \pi_j(z). \quad (\text{C.23})$$

The Laurent series expansion of $H(z)$ around the singularity z_j is

$$H(z) = \sum_{m=-\infty}^{+\infty} H_m (z - z_j)^m, \quad (\text{C.24})$$

and its principal part is given by

$$\pi_j(z) = \sum_{m=-\infty}^{-1} H_m (z - z_j)^m. \quad (\text{C.25})$$

To compute the principal part at z_j , first compute the constant $\hat{\pi}_j$ as

$$\hat{\pi}_j = \lim_{z \rightarrow z_j} (z - z_j) H(z), \quad (\text{C.26})$$

and then set

$$\pi_j(z) = \hat{\pi}_j (z - z_j). \quad (\text{C.27})$$

In this section, we first establish notation and introduce relevant mathematical definitions. We then formalize the notion of confounding dynamics, presenting a simple example that shows the mechanism at work.

C.8 FREQUENCY DOMAIN AND CONFOUNDING DYNAMICS We conclude this primer with a general discussion of how the results presented above are useful in the analysis of equilibria with incomplete information, and how they greatly facilitate the study of equilibria with confounding dynamics. Throughout the paper, we work in the space of polynomials in the lag operator L with square-summable coefficients that operate on Gaussian random variables. In our framework, any stochastic process ω_t can always be written as

$$\omega_t = Q(L)\varepsilon_t = \sum_{j=0}^{\infty} Q_j L^j \varepsilon_t, \quad (\text{C.28})$$

where $\sum_{j=0}^{\infty} |Q_j|^2 < \infty$, and $\varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon)$, are innovations identically and independently distributed over time. In linear-Gaussian environments, working with representations of the form of (C.28), and their functional equivalents,

²⁷Our methodology can also handle unit roots. For example, suppose $\theta_t = (1 - L)A(L)\varepsilon_t$. The solution, $Q(L)\varepsilon_t$, would then inherit the unit root via the cross-equation restriction.

²⁸As discussed by Whiteman (1983), the problem remains even if $\zeta = \tilde{\zeta}$.

has three advantages for analyzing rational expectations models with incomplete information.²⁹

First, representation (C.28) is general in the sense that it can accommodate both autoregressive (AR) and moving-average (MA) components of any order. This is especially useful when searching for an equilibrium because it avoids the need to specify a conjecture with a specific ARMA order. Regardless of the complexity of the equilibrium conditions that emerge in models of dispersed information (e.g., infinite regress in expectations), the solution will take the form of (C.28).³⁰

Second, the Wold Representation Theorem ensures that processes like ω_t can always be written *uniquely* as a linear combination of a moving average representation where the innovations are the linear forecast errors for ω_t , conditional on any linear-Gaussian information set [Brockwell and Davis (1987)]. That is, the Wold representation establishes the invertibility of $Q(L)$ and one may write $Q(L)^{-1}\omega_t = \varepsilon_t$, which implies that the space spanned by $\{\omega_t, \omega_{t-1}, \dots\}$ is equivalent (in mean-square norm) to the space spanned by $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$. Consequently, one can apply the optimal prediction formulas derived by Wiener-Kolmogorov [Whittle (1983)] to compute the conditional expectation of processes like ω_t .

Third, the Riesz-Fischer Theorem [see Sargent (1987)] establishes an isometric, isomorphic mapping from the space of lag polynomials with square-summable coefficients $Q(L)$ to the space of analytic complex-valued functions, where (C.28) is represented as $Q(z)$, but with $z \in \mathbb{C}$. In several key steps of the analysis in this paper we find it convenient to exploit the properties of such functions, which allows us to derive simple existence and uniqueness conditions for rational expectations equilibria with incomplete and dispersed information following Whiteman (1983). In a slight abuse of notation, we employ L and z interchangeably when working in the space of analytic functions.

We now restrict our focus to the formulation of optimal prediction formulas, which is where confounding dynamics emerge.

Suppose that we would like to formulate the prediction of ω_{t+j} so as to minimize the mean-squared forecast error, conditional on the observation of the history of a $n \times 1$ vector of variables, s_t , up to time t . To denote such history, we use the compact notation, $s^t \equiv \{s_{t-j}\}_{j=0}^{\infty}$. Let

$$s_t = \Gamma(L)u_t, \quad (\text{C.29})$$

where $\Gamma(L)$ is an $n \times m$ matrix with each element being a square-summable lag polynomial in non-negative powers of L , and u_t an $m \times 1$ vector of i.i.d. Gaussian shocks with variance-covariance normalized to the $m \times m$ identity matrix. Let $g_{ss}(z)$ be the variance-covariance generating function for the process s_t ,³¹ then one has that $g_{ss}(z) = \Gamma(z)\Gamma(z^{-1})^\top$. Similarly, one can define the covariance generating function between the joint processes ω_t and s_t , which is given by $g_{\omega s}(z) = Q(z)\sigma_\varepsilon\Gamma(z^{-1})^\top$. The prediction for ω_{t+j} that minimizes the mean squared forecast error corresponds to a linear combination of current and past realizations of s_t , denoted by

$$\mathcal{P}(\omega_{t+j}|s^t) = \Pi(L)s_t. \quad (\text{C.30})$$

Here, $\Pi(L)$ is a $1 \times n$ vector of square-summable lag polynomials in non-negative powers of L , whose form is provided by the Wiener-Kolmogorov formula

$$\Pi(L) = [L^{-j}g_{\omega s}(L)(\Gamma^*(L^{-1})^\top)^{-1}]_+ \Gamma^*(L)^{-1}. \quad (\text{C.31})$$

Expression (C.31) has several moving parts that require some unpacking. Let us first consider the special case $s_t = \omega_t$, which implies $\Gamma(L) = Q(L)$ and $u_t = \varepsilon_t$, so that the prediction problem is one in which we would like to predict the future realizations ω_{t+j} , for $j \geq 1$, using its own past, ω^t . Given the form of ω_t in (C.28), the best prediction is one

²⁹We are not the first to highlight these advantages [see, Hansen and Sargent (1980), Futia (1981), Townsend (1983b), Kasa (2000)].

³⁰Townsend (1983a) elaborates extensively on the advantages of using representations such as (C.28) when solving for equilibria that harbor an infinite regress in expectations. In Section 3.5 we show the complex form that the infinite regress in expectations takes in our framework.

³¹Recall from above that the variance-covariance generating function of a stationary Gaussian process is defined as the Fourier transform of its correlogram, which is the collection of covariances at all horizons.

that carries $Q(L)$, j periods forward and uses the best estimates of the infinite history of innovations $\{\varepsilon_{t+j}, \varepsilon_{t+j-1}, \dots\}$, to compute ω_{t+j} . The best estimates of $\{\varepsilon_{t+j}, \varepsilon_{t+j-1}, \dots, \varepsilon_{t+1}\}$ are clearly equal to the unconditional average, 0. It follows that they should not appear in the optimal prediction. This is achieved by the operator $[\]_+$ in (C.31), known as the “annihilating operator”, which instructs us to ignore the first j coefficients of its argument.

The best estimates for $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$, on the other hand, are usually not zero, and correspond to the innovations in the Wold fundamental representation for ω_t . The decomposition of the information set into its Wold fundamental representation is achieved by $\Gamma^*(L)$ in (C.31). Whether $\Gamma^*(L)$ is equal to $\Gamma(L)$ depends on the invertibility properties of the analytic function $\Gamma(z)$. For our special case, if $Q(z)$ is invertible for z inside the unit circle, then the Wold fundamental representation corresponds exactly with the history $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$, and the prediction formula $\mathcal{P}(\omega_{t+j}|\omega^t) = \tilde{\Pi}(L)\omega_t$, has

$$\tilde{\Pi}(L) = [L^{-j}Q(L)]_+Q(L)^{-1}. \quad (\text{C.32})$$

Note that all the steps just described are clearly at work here: L^{-j} carries the function $Q(L)$ j periods forward; the operator $[\]_+$ sets the estimates of innovations from $t+1$ to $t+j$ to zero, by annihilating Q_0, Q_1, \dots, Q_{j-1} ; finally, $Q(L)^{-1}$ makes sure that once $\Pi(L)$ is multiplied by ω_t , the Wold fundamental innovations $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ result.

Expression (C.32) is a special case of (C.31) because, in general, $\Gamma(z)$ is not invertible for z inside the unit circle, and solving for $\Gamma^*(z)$ is at the core of the solution to the prediction problem. Formally, $\Gamma^*(z)$ corresponds to the “canonical” factorization of the variance-covariance generating function $g_{ss}(z)$, so that

$$g_{ss}(z) = \Gamma^*(z)\Gamma^*(z^{-1})^\top. \quad (\text{C.33})$$

The canonical factorization answers the question: What space is spanned by the observables s^t ? If $\Gamma(z)$ is invertible in non-negative powers of z , then s^t will span the innovations u^t . If $\Gamma(z)$ is non-invertible, then $\Gamma^*(z)$ must be determined and it will span a space that is strictly smaller than that spanned by u^t . The existence of $\Gamma^*(z)$ is guaranteed by the Wold (fundamental) Representation Theorem. The computation of $\Gamma^*(z)$ can be quite involved, however, and there exist several methods for achieving it. In our setting, we follow the steps outlined in Rozanov (1967).³²

The most straightforward reason for $\Gamma(z)$ to be non-invertible for prediction purposes, is when $m > n$. In such case, the dimension of the vector of underlying shocks, u_t , is greater than the vector of signals, s_t , at any t . This source of non-invertibility is what is typically assumed in the incomplete information literature. A different, and often more subtle reason, is when the elements in $\Gamma(z)$ combine in a way to create an “internal” source of non-invertibility. While the analysis of Section 2 formally illustrates this claim, the easiest way to see this is to consider the case of $m = n$, so that $\Gamma(z)$ is a square matrix. $\Gamma(z)$ is non-invertible for prediction purposes if its determinant vanishes at one or more points inside the unit circle. In this case, despite having the same number of shocks and signals, non-invertibility stems from the way signals combine over time and become themselves a source of noise.

We are interested in this internal source of non-invertibility, which we term “Confounding Dynamics”. Confounding dynamics naturally arise also when $m > n$, in which case they compound with the first source of non-invertibility, but their characterization can be substantially more involved. Because our focus is on showing that confounding

³²Whittle (1983) shows that the computation of $\Gamma^*(L)$ is the dual to the solution of the n^{th} order Ricatti equation in the Kalman filter approach to optimal linear-quadratic prediction. In both cases the objective is to figure out the variance-covariance of the optimal prediction errors. Hence, *if* the Ricatti equation is solved, then $\Gamma^*(z)$ can be determined; and *if* $\Gamma^*(L)$ is computed, then a solution to the Ricatti equation can be determined. In this paper we take the latter approach, while other authors, most notably Huo and Takayama (2016), take the former. Whether it is easier to solve the Ricatti equation or to obtain the canonical factorization often depends on the application at hand, and whether one is looking for an analytical or a numerical solution. The canonical factorization, however, has a wider scope of applicability. Whittle (1983) presents a generalized approach – which he terms the Hamiltonian approach – where the prediction problem is solved by the canonical factorization of a matrix function obtained by augmenting the plant and observation equations of a Kalman system with the saddle point conditions of an Hamiltonian optimization problem. Huertgen, Hoffmann, Rondina, and Walker (2016) make an explicit connection between these solution methodologies.

dynamics can endogenously emerge in equilibrium, in the paper we focus mostly on the $m = n$ case. In Appendix [B.8](#) we analyze an example where the two sources of non-invertibility are simultaneously present, while we provide a more formal treatment of the $m > n$ case in Appendix [B.3](#). The following definition formalizes the above discussion.