Econometric Theory, **42**, 2026, 135–191.
doi:10.1017/S026646662400029X

THE SPECTRAL APPROACH TO LINEAR RATIONAL EXPECTATIONS MODELS

MAJID M. AL-SADOON
Durham University Business School

This paper considers linear rational expectations models in the frequency domain. The paper characterizes existence and uniqueness of solutions to particular as well as generic systems. The set of all solutions to a given system is shown to be a finite-dimensional affine space in the frequency domain. It is demonstrated that solutions can be discontinuous with respect to the parameters of the models in the context of nonuniqueness, invalidating mainstream frequentist and Bayesian methods. The ill-posedness of the problem motivates regularized solutions with theoretically guaranteed uniqueness, continuity, and even differentiability properties.

1. INTRODUCTION

Spectral (or frequency-domain) analysis of covariance stationary processes is a cornerstone of time series analysis. Since its beginnings in the late 1930s, it has benefited from being at the intersection of a number of fundamental mathematical subjects including probability, functional analysis, and complex analysis (Rozanov, 1967; Nikolski, 2002; Bingham, 2012a, 2012b). Almost concurrently, economic theory began to focus on expectations of future earnings, prices, interest rates, and so forth as determinants of present economic activity (Knight, 1921; Keynes, 1936; Cagan, 1956). This led to the pioneering work of Muth (1961), who proposed that expectations, rather than being arbitrary inputs into models or arbitrarily determined within models, could be made both endogenous and model-consistent, hence "rational" (see Pesaran, 1987 for further historical context). Today, rational expectations models are the mainstay of business cycle research (Canova, 2011; DeJong and Dave, 2011; Herbst and Schorfheide, 2016). This paper attempts to bridge the gap between the two strands of literature.

Classical linear systems, such as vector autoregressive moving average (VARMA) models, are linear transformations from an input process to an output process, with present values of the output depending linearly on present and past values of the input as well as past values of the output in every time period.

Thanks are due to Todd Walker, Bernd Funovits, Mauro Bambi, Piotr Zwiernik, Abderrahim Taamouti, Benedikt Pötscher, three anonymous referees, and seminar participants at Heriot-Watt University, Universitat Pompeu Fabra, University of Bologna, and Aarhus University. Address correspondence to Majid M. Al-Sadoon, Department of Economics, Durham University Business School, Durham, UK; e-mail: majid.al-sadoon@durham.ac.uk.

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

The linear rational expectations model (LREM) class extends classical linear systems by allowing linear dependence on expectations of future values of the output as well. Such models arise naturally from the inter-temporal optimization problems of households and firms in economic modeling. Spectral analysis has focused almost entirely on classical linear systems (Brockwell and Davis, 1991; Pourahmadi, 2001; Lindquist and Picci, 2015). Although a number of works have considered LREMs in the frequency domain (Whiteman, 1983; Onatski, 2006; Tan and Walker, 2015; Tan, 2019; Meyer-Gohde and Tzaawa-Krenzler, 2023), none have attempted a general account that parallels the aforementioned textbook treatments. Thus, the first aim of this paper is to situate LREMs in the frequency-domain literature at a high level of generality that extends the aforementioned textbook treatments.

Using the spectral representation of covariance stationary processes due to Kolmogorov (1939, 1941a, 1941b) and Cramér (1940, 1942), this paper recasts the LREM problem in the classical Hilbert space of the frequency-domain literature. This is demonstrated concretely on simple scalar models before generalizing to multivariate models. Whereas classical spectral analysis focused entirely on the backward shift operator, the new spectral analysis of LREMs requires the introduction of a new operator associated with expectations. The LREM problem then reduces to a linear system in Hilbert space. However, unlike the special case of VARMA models, LREMs cannot be solved by simply inverting a polynomial matrix due to the presence of the new expectation operators. Solving the LREM problem requires using the method of Wiener–Hopf factorization (Wiener and Hopf, 1931; Gohberg and Fel'dman, 1974; Clancey and Gohberg, 1981). This paper characterizes existence and uniqueness of solutions to particular as well as generic LREMs, generalizing results by Onatski (2006). The set of all solutions to a given LREM is shown to be a finite-dimensional affine space in the frequency domain. The dimension of this space is expressed much more simply than in Funovits (2017, 2020). It is important to note that the underlying assumptions in this paper are weaker than in the previous literature (Whiteman, 1983; Onatski, 2006; Tan and Walker, 2015; Tan, 2019; Meyer-Gohde and Tzaawa-Krenzler, 2023), which requires the exogenous process to have a purely nondeterministic Wold (1938) representation, an assumption that is demonstrably unnecessary. The weaker assumptions of this paper also permit a clear answer as to why unit roots must be excluded, an aspect of the theory absent from the previous literature.

The main results of the paper concern the ill-posedness of the LREM problem in macroeconometrics. Hadamard (1902) defines a problem to be well-posed if its solutions satisfy the conditions of existence, uniqueness, and continuous dependence on its parameters. The LREM problem is ill-posed because it violates not just the second condition but also the third. Indeed, it has long been accepted that nonuniqueness is a feature of the LREM problem and many techniques have been developed to select a solution for any given LREM exhibiting nonuniqueness (e.g., Taylor, 1977; McCallum, 1983; Lubik and Schorfheide, 2003; Fanelli, 2012; Farmer, Khramov, and Nicolò, 2015; Bianchi and Nicolò, 2021). This paper

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

highlights the fact that selections that have been proposed in the literature are not guaranteed to be continuous with respect to parameters of the LREM in the context of nonuniqueness. This problem seems to be either not well understood or not fully appreciated so far; to the best of the author's knowledge, Sims (2007) is the only acknowledgement of this problem.

The problem of discontinuity is quite serious because it invalidates mainstream econometric methodology as reviewed, for example, in Canova (2011), DeJong and Dave (2011), or Herbst and Schorfheide (2016). This methodology takes a black-box approach to estimation and inference, whereby data are used to construct certain objective functions (likelihood functions, Generalized Method of Moments (GMM) criterion functions, or posterior distributions), which are then either numerically optimized by the Newton–Raphson algorithm or explored by the random walk Metropolis–Hastings algorithm (or the many variations of such algorithms) to produce estimates of parameters as well as confidence or credible intervals. When the discontinuous selected solutions are fed into this methodology, the aforementioned objective functions become discontinuous, invalidating assumptions required for the aforementioned methodologies to work. This paper illustrates concretely how things can go wrong with a simple example in Section 7.

Fortunately, the literature on ill-posed problems offers an immediate solution to the problem: regularization. The idea here is that when theory is insufficient to pin down a unique solution, other information can be brought to bear. The method can be interpreted in at least two ways: (i) penalizing economically unreasonable solutions or shrinking toward economically reasonable ones or (ii) imposing prior beliefs on the frequencies of fluctuations that solutions ought to exhibit. For example, we may like to avoid solutions where certain variables vary too wildly or impose the prior that solutions to a business cycle model should exhibit fluctuations of period between 4 and 32 quarters in quarterly data. The paper provides conditions for existence and uniqueness of regularized solutions and proves that they are continuously (even differentiably) dependent on their parameters. Thus, under nonuniqueness, regularization selects solutions that can be used in any mainstream econometric method, frequentist or Bayesian.

This work is related to several recent strands in the literature. Komunjer and Ng (2011), Qu and Tkachenko (2017), Kociecki and Kolasa (2018), Al-Sadoon and Zwiernik (2019), and Kociecki and Kolasa (2023) study the identification of LREMs based on the spectral density of observables. Christiano and Vigfusson (2003), Qu and Tkachenko (2012), and Sala (2015) utilize spectral-domain methods for estimating LREMs using ideas that go back to Hansen and Sargent (1980). Chahrour and Jurado (2021) study the problem of subordination (what they call "recoverability") in the context of macroeconometric models. Al-Sadoon (2018) utilizes a generalization of Wiener–Hopf factorization in order to study unstable and nonstationary solutions of LREMs. Ephremidze, Shargorodsky, and Spitkovsky (2020) provide recent results on the continuity of spectral factorization. Finally, Al-Sadoon (2020) provides numerical algorithms for computing regularized solutions.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

This paper is organized as follows. Section 2 sets up the notation and reviews the fundamental concepts of spectral analysis of time series. Section 3 considers the solution of simple LREMs with elementary frequency-domain methods. Section 4 introduces Wiener–Hopf factorization. Section 5 sets up the LREM problem, its existence and uniqueness properties, and establishes its ill-posedness. Section 6 introduces regularized solutions. Section 7 provides an application of the theory. Section 8 concludes. Sections A–D comprise the Appendix.

2. NOTATION AND REVIEW

We will denote by $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ the sets of integers, real numbers, and complex numbers, respectively. We will denote by $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ the unit circle in $\mathbb{C}$. The letters i, j, n, and m will always stand for natural numbers. By $\mathbb{C}^{m\times n}$, we will denote the set of $m \times n$ matrices of complex numbers. We will use I_n to denote the identity $n \times n$ matrix. For $M \in \mathbb{C}^{m\times m}$, $\mathrm{tr}(M) = \sum_{i=1}^{m} M_{ii}$. For $M \in \mathbb{C}^{m\times n}$, $M^* \in \mathbb{C}^{n\times m}$ is the conjugate transpose of M, and $\|M\|_{\mathbb{C}^{m\times n}} = (\mathrm{tr}(MM^*))^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} |M_{ij}|^2\right)^{1/2}$.

All random variables in this paper are defined over a single probability space $(\Omega, \mathscr{F}, P)$. For a complex random variable x, the expectation is denoted by $\boldsymbol{E}x = \int_\Omega x(\omega)dP(\omega)$. The space $\mathscr{L}_2$ is defined as the Hilbert space of complex-valued random variables x, modulo P-almost sure equality, such that $\boldsymbol{E}|x|^2 < \infty$ with inner product and norm

$$\langle x, y\rangle = \boldsymbol{E}x\bar{y}, \qquad \|x\|^2_{\mathscr{L}_2} = \langle x, x\rangle, \qquad x, y \in \mathscr{L}_2.$$

Similar to other Hilbert spaces we will consider in this paper, $\mathscr{L}_2$ is a set of equivalence classes of functions and not a set of functions. However, mathematical convention "relegate[s] this distinction to the status of a tacit understanding" (Rudin, 1986, p. 67). Thus, we will write "$x = y$" instead of "$x(\omega) = y(\omega)$ for P-almost all $\omega \in \Omega$" and similarly for elements of all of the other Hilbert spaces we consider in this paper. The Hilbert space $\mathscr{L}_2^n$ is defined as the n-fold Cartesian product, $\mathscr{L}_2 \times \cdots \times \mathscr{L}_2$, consisting of column vectors $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$, $x_j \in \mathscr{L}_2$, $j = 1, \ldots, n$, with the inner product and norm

$$\langle\langle x, y\rangle\rangle = \sum_{j=1}^{n} \langle x_j, y_j\rangle, \qquad \|x\|^2_{\mathscr{L}_2^n} = \langle\langle x, x\rangle\rangle, \qquad x, y \in \mathscr{L}_2^n.$$

If $\mathscr{S} \subset \mathscr{L}_2$ is a closed subspace and $x \in \mathscr{L}_2^n$, then the minimum of

$$\|x - y\|^2_{\mathscr{L}_2^n} = \sum_{j=1}^{n} \|x_j - y_j\|^2_{\mathscr{L}_2},$$

with respect to $y \in \mathscr{S}^n = \mathscr{S} \times \cdots \times \mathscr{S}$, is attained by minimizing each term on the right-hand side independently with respect to $y_j \in \mathscr{S}$, $j = 1, \ldots, n$. It follows

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

that the orthogonal projection of $x \in \mathscr{L}_2^n$ onto $\mathscr{S}^n$ in $\mathscr{L}_2^n$, denoted by $\boldsymbol{P}(x|\mathscr{S}^n)$, is given by $\begin{bmatrix} \boldsymbol{P}(x_1|\mathscr{S}) \\ \vdots \\ \boldsymbol{P}(x_n|\mathscr{S}) \end{bmatrix}$, where $\boldsymbol{P}(x_j|\mathscr{S})$ is the orthogonal projection of x_j onto $\mathscr{S}$ in $\mathscr{L}_2, j = 1, \ldots, n$.

Let $\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$ be a doubly infinite sequence in $\mathscr{L}_2^n$. We will refer to this stochastic process simply as ξ. The jth element of ξ_t will be denoted by ξ_{jt}. If, for all $t, s \in \mathbb{Z}$, $\boldsymbol{E}\xi_t = \boldsymbol{E}\xi_s$ and $\boldsymbol{E}\xi_t\xi_s^*$ depends on t and s only through $t - s$, we say that ξ is a covariance stationary process. Given ξ, we may define a number of useful objects.

Let $\mathscr{H}$ be the closure in $\mathscr{L}_2$ of the set of all finite complex linear combinations of the set $\{\xi_{jt} : j = 1, \ldots, n, t \in \mathbb{Z}\}$. We define $\mathscr{H}^n \subset \mathscr{L}_2^n$ to be the n-fold Cartesian product of $\mathscr{H}$.

Traditionally, spectral analysis utilizes the forward (rather than backward) shift operator,

$$\boldsymbol{U} : \xi_{jt} \mapsto \xi_{jt+1}, \qquad j = 1, \ldots, n, \quad t \in \mathbb{Z}.$$

This operator extends uniquely to a unitary operator on $\mathscr{H}$ (Rozanov, 1967, Thm. I.4.1).

We may next define the unique spectral measure F on the unit circle, $\mathbb{T}$, which satisfies

$$\boldsymbol{E}\xi_t\xi_s^* = \int z^{t-s} dF, \qquad t, s \in \mathbb{Z}$$

(Rozanov, 1967, Thms. I.5.1 and I.5.2). Many textbooks express the integral above as $\int_{-\pi}^{\pi} e^{\mathrm{i}(t-s)\lambda} dF(\lambda)$; the notation here is substantially more compact. When each element of F is absolutely continuous with respect to normalized Lebesgue measure on $\mathbb{T}$, defined by

$$\mu\left(\{e^{\mathrm{i}\lambda} : a \le \lambda \le b\}\right) = \frac{1}{2\pi}(b - a), \qquad 0 \le b - a \le 2\pi,$$

then the Radon–Nikodým derivative $dF/d\mu$ is the spectral density matrix of ξ. Note that the spectral density of n-dimensional standardized white noise is I_n. Another important example of a spectral measure is the Dirac measure at $w \in \mathbb{T}$ defined as

$$\delta_w(\Lambda) = \begin{cases} 1, & w \in \Lambda, \\ 0, & w \notin \Lambda, \end{cases}$$

for Borel subsets $\Lambda \subset \mathbb{T}$. Note that $F = \delta_w I_n$ is the spectral measure of the purely deterministic process $\xi_t = \xi_0 w^t$ for $t \in \mathbb{Z}$, with $\boldsymbol{E}\xi_0\xi_0^* = I_n$.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

We may also define the Hilbert space H of all Borel-measurable mappings $\phi : \mathbb{T} \to \mathbb{C}^{1\times n}$ (i.e., $\phi(z)$ is a row vector for $z \in \mathbb{T}$) such that $\int \phi dF \phi^* < \infty$ with inner product and norm

$$(\phi, \varphi) = \int \phi dF \varphi^*, \qquad \|\phi\|_H^2 = (\phi, \phi), \qquad \phi, \varphi \in H,$$

where we identify $\phi = \varphi$ if $\|\phi - \varphi\|_H^2 = \int (\phi - \varphi) dF (\phi - \varphi)^* = 0$ (Rozanov, 1967, Lem. I.7.1). We define the Hilbert space $H^m = H \times \cdots \times H$ to be the m-fold Cartesian product of H consisting of the $m \times n$ matrices $\phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_m \end{bmatrix}$, $\phi_j \in H$, $j = 1, \ldots, m$ endowed with the inner product and norm

$$((\phi, \varphi)) = \sum_{j=1}^{m} (\phi_j, \varphi_j), \qquad \|\phi\|_{H^m}^2 = ((\phi, \phi)), \qquad \phi, \varphi \in H^m.$$

By the same argument as used earlier, if $S \subset H$ is a closed subspace, the orthogonal projection of $\phi \in H^m$ onto S^m in H^m, denoted by $\boldsymbol{P}(\phi | S^m)$, is given by $\begin{bmatrix} \boldsymbol{P}(\phi_1 | S) \\ \vdots \\ \boldsymbol{P}(\phi_m | S) \end{bmatrix}$, where $\boldsymbol{P}(\phi_j | S)$ is the orthogonal projection of ϕ_j onto S in H, $j = 1, \ldots, m$.

The spectral representation theorem states that

$$\xi_t = \int z^t d\Phi, \qquad t \in \mathbb{Z},$$

for a column vector of random measures Φ with $\Phi(\Lambda) \in \mathscr{H}^n$ and $\boldsymbol{E}\Phi(\Lambda)\Phi^*(\Lambda) = F(\Lambda)$ for Borel subsets $\Lambda \subset \mathbb{T}$ (Rozanov, 1967, Thm. I.4.2). Many textbooks express the integral above equivalently as $\int_{-\pi}^{\pi} e^{\mathrm{i}t\lambda} dZ(\lambda)$, where $Z(\lambda) = \Phi(\{e^{\mathrm{i}\tau} : \tau \in (-\pi, \lambda]\})$ for $\lambda \in (-\pi, \pi]$; again, the notation here is substantially more compact. The spectral representation theorem establishes a unitary mapping $H^m \to \mathscr{H}^m$ defined by $\phi \mapsto h = \int \phi d\Phi$ (Rozanov, 1967, p. 32). We call ϕ the spectral characteristic of h. Thus,

$$\left\langle \left\langle \int \phi d\Phi, \int \varphi d\Phi \right\rangle \right\rangle = ((\phi, \varphi)), \qquad \left\| \int \phi d\Phi \right\|_{\mathscr{L}_2^m}^2 = \|\phi\|_{H^m}^2, \qquad \phi, \varphi \in H^m.$$

Denote by $e_j \in \mathbb{C}^{1\times n}$ the row vector whose elements are all equal to zero except for the jth, which is equal to one. The spectral representation theorem implies that $\mathscr{H}_t$, the closure in $\mathscr{L}_2$ of the set of all finite complex linear combinations of $\{\xi_{js} : j = 1, \ldots, n, s \le t\}$, is in correspondence with the closure of the set of all finite complex linear combinations of $\{z^s e_j : j = 1, \ldots, n, s \le t\}$ in H, denoted H_t. Let $\mathscr{H}_t^m$ and H_t^m be m-fold Cartesian products of $\mathscr{H}_t$ and H_t, respectively. This implies that

$$\boldsymbol{P}\left(\int \phi d\Phi \,\middle|\, \mathscr{H}_t^m \right) = \int \boldsymbol{P}(\phi | H_t^m) d\Phi, \qquad t \in \mathbb{Z}.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Thus, the best linear prediction of ξ_{t+s} in terms of $\xi_t, \xi_{t-1}, \ldots$ is given by

$$\boldsymbol{P}(\xi_{t+s}|\mathscr{H}_t^n) = \int \boldsymbol{P}(z^{t+s}I_n|H_t^n)d\Phi, \qquad s,t \in \mathbb{Z}.$$

It is easily established that $\mathscr{H}_t = \boldsymbol{U}^t\mathscr{H}_0$ and $H_t = z^tH_0$ for all $t \in \mathbb{Z}$ so that $\boldsymbol{P}(\boldsymbol{U}^t\xi|\mathscr{H}_t) = \boldsymbol{U}^t\boldsymbol{P}(h|\mathscr{H}_0)$ for all $t \in \mathbb{Z}$ and $h \in \mathscr{H}$ and likewise $\boldsymbol{P}(z^t\phi|H_t) = z^t\boldsymbol{P}(\phi|H_0)$ for all $t \in \mathbb{Z}$ and $\phi \in H$ (Rozanov, 1967, p. 52). Thus,

$$\boldsymbol{P}(\xi_{t+s}|\mathscr{H}_t^n) = \int z^t\boldsymbol{P}\left(z^sI_n|H_0^n\right)d\Phi, \qquad s,t \in \mathbb{Z}.$$

If ζ is an additional covariance stationary processes such that for every $t,s \in \mathbb{Z}$, $\boldsymbol{E}\zeta_s\xi_t^*$ depends on t and s only through $t-s$, we will say that ζ is causal in ξ if $\zeta_0 \in \mathscr{H}_0^m$. Finally, if ν is causal in ξ and satisfies $\boldsymbol{P}(\nu_{t+1}|\mathscr{H}_t) = 0$ for all $t \in \mathbb{Z}$, we call ν an innovation process.

3. EXAMPLES

Armed with the basic machinery above, we now make a first attempt at solving LREMs in the frequency domain. We will see that solving simple univariate LREMs involves only elementary spectral-domain techniques as discussed in textbook treatments of spectral analysis such as Brockwell and Davis (1991) or Pourahmadi (2001). The methods also provide strong hints to the general approach to solving LREMs. In this section, ξ is a scalar covariance stationary process with $\boldsymbol{U}$, Φ, F, $\mathscr{H}$, and H defined as in the previous section.

3.1. The Autoregressive Model

We begin on familiar territory with the stationary autoregression

$$X_t - \alpha X_{t-1} = \xi_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where $|\alpha| < 1$. The frequency-domain analysis of this model is available in many textbooks. For completeness, we provide a treatment here that is geared toward understanding the more general cases to come.

We require a covariance stationary solution causal in ξ (further motivation of this restriction can be found in Section 5). Thus, we require

$$X_t = \int z^t\phi d\Phi, \qquad t \in \mathbb{Z},$$

for some spectral characteristic $\phi \in H_0$.

Notice that we may restrict attention to the equation

$$(1 - \alpha\boldsymbol{U}^{-1})X_0 = \xi_0.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

If a solution X_0 exists, then the rest of the process can be generated as

$$X_t = \boldsymbol{U}^t X_0, \qquad t \in \mathbb{Z},$$

and clearly satisfies (1).

Thus, we must solve for ϕ in

$$\int (1 - \alpha z^{-1}) \phi d\Phi = \int d\Phi.$$

Since integration with respect to Φ is a unitary mapping $H \to \mathscr{H}$, it has an inverse and the equation above is equivalent to

$$(1 - \alpha z^{-1})\phi = 1 \tag{2}$$

in the frequency domain.

The linear mapping $\phi \mapsto \alpha z^{-1}\phi$ on H_0 is bounded in norm by $|\alpha| < 1$; thus, the mapping $\phi \mapsto (1 - \alpha z^{-1})\phi$ is invertible and

$$\phi = \sum_{s=0}^{\infty} \alpha^s z^{-s}$$

is the unique solution to (2), where the summation is understood to converge with respect to H-norm (Gohberg, Goldberg, and Kaashoek, 2003a, Thm. 2.8.1). Thus, we arrive at the unique solution to X_0,

$$X_0 = \int \sum_{s=0}^{\infty} \alpha^s z^{-s} d\Phi = \sum_{s=0}^{\infty} \alpha^s \int z^{-s} d\Phi = \sum_{s=0}^{\infty} \alpha^s \xi_{-s}.$$

The interchange of the summation and the stochastic integral is admissible because the stochastic integral is a bounded linear operator mapping from H to $\mathscr{H}$ and the inner summation converges in H. It follows that the unique covariance stationary solution is

$$X_t = \boldsymbol{U}^t X_0 = \sum_{s=0}^{\infty} \alpha^s \boldsymbol{U}^t \xi_{-s} = \sum_{s=0}^{\infty} \alpha^s \xi_{t-s}, \qquad t \in \mathbb{Z}.$$

The operator $\boldsymbol{U}^t$ is interchangeable with the summation because it is a bounded linear operator on $\mathscr{H}$ and the summation converges in $\mathscr{L}_2$.

3.2. The Cagan Model

The Cagan model is given as

$$X_t - \beta \boldsymbol{P}(X_{t+1} | \mathscr{H}_t) = \xi_t, \qquad t \in \mathbb{Z}, \tag{3}$$

with $|\beta| < 1$. Again, we look for a covariance stationary solution causal in ξ and we restrict attention to the equation

$$X_0 - \beta \boldsymbol{P}(UX_0 | \mathscr{H}_0) = \xi_0,$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

and, following a similar argument to that used above, we arrive at the underlying frequency-domain problem,

$$\phi - \beta \boldsymbol{P}(z\phi|H_0) = 1. \tag{4}$$

Since the linear mapping $\phi \mapsto \beta \boldsymbol{P}(z\phi|H_0)$ on H_0 is bounded in norm by $|\beta| < 1$, we see that the mapping $\phi \mapsto \phi - \beta \boldsymbol{P}(z\phi|H_0)$ is invertible so that

$$\phi = \sum_{s=0}^{\infty} \beta^s \boldsymbol{P}(z^s|H_0)$$

is the unique solution to (4), where the summation is understood to converge with respect to H-norm (Gohberg et al., 2003a, Thm. 2.8.1). Thus, we arrive at the unique solution to X_0,

$$X_0 = \int \sum_{s=0}^{\infty} \beta^s \boldsymbol{P}(z^s|H_0) d\Phi = \sum_{s=0}^{\infty} \beta^s \int \boldsymbol{P}(z^s|H_0) d\Phi = \sum_{s=0}^{\infty} \beta^s \boldsymbol{P}(\xi_s|\mathscr{H}_0).$$

The interchange of the summation and the stochastic integral is admissible because the stochastic integral is a bounded linear operator mapping H to $\mathscr{H}$ and the inner summation converges in H. It follows that the unique covariance stationary solution is

$$X_t = \boldsymbol{U}^t X_0 = \sum_{s=0}^{\infty} \beta^s \boldsymbol{U}^t \boldsymbol{P}(\xi_s|\mathscr{H}_0) = \sum_{s=0}^{\infty} \beta^s \boldsymbol{P}(\xi_{t+s}|\mathscr{H}_t), \qquad t \in \mathbb{Z}.$$

The operator $\boldsymbol{U}^t$ is interchangeable with the summation because it is a bounded linear operator on $\mathscr{H}$ and the summation converges in $\mathscr{L}_2$.

3.3. The Mixed Model

Now suppose we have the more general model

$$a\boldsymbol{P}(X_{t+1}|\mathscr{H}_t) + bX_t + cX_{t-1} = \xi_t, \qquad t \in \mathbb{Z}, \tag{5}$$

where $a, b, c \in \mathbb{C}$ and $ac \neq 0$. This leads to the frequency-domain equation

$$\boldsymbol{P}((az + b + cz^{-1})\phi|H_0) = 1. \tag{6}$$

As noted by Sargent (1979), the solution to this system depends on the factorization of

$$M(z) = az + b + cz^{-1} = az^{-1}(z - \delta)(z - \gamma).$$

We assume, without loss of generality, that $|\gamma| \leq |\delta|$. There are four cases to consider.

Suppose $|\gamma| < 1 < |\delta|$. We can then write $M(z) = a(z - \delta)(1 - \gamma z^{-1})$ and express the system as

$$\boldsymbol{P}((1 - \delta^{-1}z)\varphi|H_0) = 1, \qquad -a\delta(1 - \gamma z^{-1})\phi = \varphi.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

The first equation can be solved as in the Cagan model (since $|\delta^{-1}| < 1$), while the second can be solved as in the autoregressive model (since $|\gamma| < 1$). This procedure leads us to the unique solution

$$\phi = -\frac{1}{a\delta}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^{-s}z^{-u}\boldsymbol{P}(z^s|H_0) = -\frac{1}{a\delta}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^{-s}\boldsymbol{P}(z^{s-u}|H_{-u}).$$

In the time domain, we obtain the following solution:

$$X_0 = -\frac{1}{a\delta}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^{-s}\boldsymbol{P}(\xi_{s-u}|\mathscr{H}_{-u}),$$

which then gives us the general solution,

$$X_t = -\frac{1}{a\delta}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^{-s}\boldsymbol{P}(\xi_{s+t-u}|\mathscr{H}_{t-u}), \qquad t \in \mathbb{Z}.$$

Thus, when $|\gamma| < 1 < |\delta|$, there exists a unique solution.

Next, suppose that $|\delta| < 1$. Then we may write $M(z) = az(1-\delta z^{-1})(1-\gamma z^{-1})$ and express our system as

$$\boldsymbol{P}(z\varphi|H_0) = 1, \qquad a(1-\delta z^{-1})(1-\gamma z^{-1})\phi = \varphi.$$

The first equation does not have a unique solution in general. For example, when $F = \mu$, then $\varphi = z^{-1} + \psi$ solves the equation for any $\psi \in \mathbb{C}$. More generally, every solution is of the form

$$\varphi = z^{-1} + \psi,$$

where $\psi \in H_0$ with $z\psi$ orthogonal to H_0. We can then solve for ϕ as

$$\phi = \frac{1}{a}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^s z^{-s-u}(z^{-1} + \psi).$$

In the time domain, this leads to the general solution

$$X_t = \frac{1}{a}\sum_{u=0}^{\infty}\sum_{s=0}^{\infty}\gamma^u\delta^s(\xi_{t-s-u-1} + \nu_{t-s-u}), \qquad t \in \mathbb{Z},$$

where $\nu_t = \int z^t\psi\, d\Phi \in \mathscr{H}_t$ for $t \in \mathbb{Z}$ is an arbitrary innovation process. Indeed, $\nu_t \in \mathscr{H}_t$ for all $t \in \mathbb{Z}$ because $\psi \in H_0$ and ν_{t+1} is orthogonal to $\mathscr{H}_t$ for all $t \in \mathbb{Z}$ because $\boldsymbol{P}(\nu_{t+1}|\mathscr{H}_t) = \int \boldsymbol{P}(z^{t+1}\psi|H_t)d\Phi = \int z^t\boldsymbol{P}(z\psi|H_0)d\Phi = 0$ as $z\psi$ is orthogonal to H_0. Thus, when $|\delta| < 1$, there exist potentially infinitely many solutions.

Now suppose $|\gamma| > 1$. Then we may write $M(z) = a\delta\gamma z^{-1}(1-\delta^{-1}z)(1-\gamma^{-1}z)$ and

$$\boldsymbol{P}(a\delta\gamma(1-\delta^{-1}z)(1-\gamma^{-1}z)\varphi|H_0) = 1, \qquad z^{-1}\phi = \varphi.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Clearly, φ can be solved as in the Cagan model to produce

$$\varphi = \frac{1}{a\delta\gamma} \sum_{u=0}^{\infty} \sum_{s=0}^{\infty} \gamma^{-u} \delta^{-s} \boldsymbol{P}(z^{s+u}|H_0).$$

This implies that

$$z^{-1}\phi = \frac{1}{a\delta\gamma} \sum_{u=0}^{\infty} \sum_{s=0}^{\infty} \gamma^{-u} \delta^{-s} \boldsymbol{P}(z^{s+u}|H_0).$$

However, this equation cannot hold generally. To see this, let ξ be a standard white noise so that $F = \mu$. Then $\{z^s : s \in \mathbb{Z}\}$ is an orthonormal set and the equation above reduces to

$$z^{-1}\phi = \frac{1}{a\delta\gamma}.$$

Since ϕ is in the closure of the linear span of $\{z^s : s \leq 0\}$, the zeroth Fourier coefficient of the left-hand side is equal to zero, while that of the right-hand side is equal to $\frac{1}{a\delta\gamma} \neq 0$. So there is no solution in general when $|\gamma| > 1$.

Finally, suppose either $|\gamma| = 1$ or $|\delta| = 1$ so that $M(w) = 0$ for some $w \in \mathbb{T}$. Then there is no solution in general, in the sense that there exist processes ξ for which no covariance stationary solution X can be found. To see this, let $F = \delta_w$, the Dirac measure at w. If a solution $\phi \in H_0$ to (6) exists, then it must satisfy

$$\|\phi\|_{H_0}^2 = \int |\phi|^2 dF = |\phi(w)|^2 < \infty.$$

This then implies that $M\phi = M(w)\phi(w) = 0$ in H, which implies that $\boldsymbol{P}(M\phi|H_0) = 0$, contradicting the fact that $\boldsymbol{P}(M\phi|H_0) = 1$. Thus, when $|\gamma| = 1$ or $|\delta| = 1$, there exists no solution in general.

We will see that the analysis of the simple models above contains most of the elements necessary for the general multivariate case with more than one lead and/or lag and arbitrary covariance stationary ξ. The reader interested in a more complete elementary analysis of the simple models above (e.g., considering the case $|\beta| > 1$ of the Cagan model) is directed to previous versions of this paper available on arXiv.

4. WIENER–HOPF FACTORIZATION

The approach we have taken in the last section is well understood in the theory of convolution equations (Gohberg and Fel'dman, 1974). The requisite factorization of $M(z)$ into two parts, a part to solve like the Cagan model and a part to solve like the autoregressive model, is known as a Wiener–Hopf factorization (Wiener and Hopf, 1931). In this section, we state the basic concepts and properties of Wiener–Hopf factorization that we will need.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

DEFINITION 1. *Let $\mathcal{W}$ be the class of functions $M : \mathbb{T} \to \mathbb{C}$ defined by*

$$M(z) = \sum_{s=-\infty}^{\infty} M_s z^s, \tag{7}$$

where $M_s \in \mathbb{C}$ for $s \in \mathbb{Z}$ and $\sum_{s=-\infty}^{\infty} |M_s| < \infty$. Define $\mathcal{W}_{\pm} \subset \mathcal{W}$ to be the class of functions (7) with $M_s = 0$ for $s \lessgtr 0$. The sets $\mathcal{W}^{m\times n}$, $\mathcal{W}_{\pm}^{m\times n}$ are defined as the sets of matrices of size $m \times n$ populated by elements of $\mathcal{W}$ and $\mathcal{W}_{\pm}$, respectively.

The class of functions $\mathcal{W}$ is known as the Wiener algebra in the functional analysis literature (Gohberg, Goldberg, and Kaashoek, 1993, Sect. XXIX.2). Note that every function analytic in a neighborhood of $\mathbb{T}$ defines an element of $\mathcal{W}$, but the opposite inclusion does not hold (e.g., $\sum_{s=0}^{\infty} s^{-2} z^s \in \mathcal{W}_+$ diverges outside $\mathbb{T}$). For $M \in \mathcal{W}^{m\times n}$, we also have that

$$\|M\|_{\infty} \leq \sum_{s=-\infty}^{\infty} \|M_s\|_{\mathbb{C}^{m\times n}} < \infty,$$

where $\|M\|_{\infty}$ is the μ-essential supremum of $\|M(z)\|_{\mathbb{C}^{m\times n}}$. We will also need the operator

$$\left[\sum_{s=-\infty}^{\infty} M_s z^s\right]_- = \sum_{s=-\infty}^{0} M_s z^s$$

defined over $\mathcal{W}^{m\times n}$ and, more generally, over the class of square summable series satisfying $\sum_{s=-\infty}^{\infty} \|M_s\|_{\mathbb{C}^{n\times m}}^2 < \infty$ (Gohberg et al., 1993, Cor. XXIII.3.3).

DEFINITION 2. *A Wiener–Hopf factorization of $M \in \mathcal{W}^{m\times m}$ is a factorization*

$$M = M_+ M_0 M_-, \tag{8}$$

where $M_+ \in \mathcal{W}_+^{m\times m}$, $M_+^{-1} \in \mathcal{W}_+^{m\times m}$, $M_- \in \mathcal{W}_-^{m\times m}$, $M_-^{-1} \in \mathcal{W}_-^{m\times m}$, M_0 is a diagonal matrix with diagonal elements $z^{\kappa_1}, \ldots, z^{\kappa_m}$, and $\kappa_1 \geq \cdots \geq \kappa_m$ are integers, called partial indices.

We remark that the factorization given in Definition 2 is termed a left factorization in the Wiener–Hopf factorization literature (Gohberg and Fel'dman, 1974, p. 185). It differs from the right factorization utilized in Onatski (2006) and Al-Sadoon (2018), where the roles of $M_{\pm}(z)$ are reversed. The difference is due to the fact that the present analysis works with the forward shift operator, whereas Onatski (2006) and Al-Sadoon (2018) work with the backward shift operator. A left factorization of M is obtained from a right factorization of M^*.

THEOREM 1. *Let $M \in \mathcal{W}^{m\times m}$ and suppose $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$. Then M has a Wiener–Hopf factorization and its partial indices are unique.*

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Proof. See Theorems VIII.1.1 and VIII.2.2 of Gohberg and Fel'dman (1974). □

The condition in Theorem 1 is minimal for existence of a Wiener–Hopf factorization and uniqueness of the M_0 part. The $M_\pm$ parts are not unique, but their general form is well understood (Gohberg and Fel'dman, 1974, Thm. VIII.1.2). The nonuniqueness of $M_\pm$ has no bearing on any of our results. Wiener–Hopf factorizations can be computed in a variety of ways (see Rogosin and Mishuris, 2016 for a recent survey).

The partial indices allow us to identify an important subset of $\mathcal{W}^{m\times m}$.

DEFINITION 3. *Let $\mathcal{W}_\circ^{m\times m}$ be the subset of $M \in \mathcal{W}^{m\times m}$ such that* $\det(M(z)) \neq 0$ *for all $z \in \mathbb{T}$ and $\kappa_1 - \kappa_m \leq 1$.*

When $m = 1$, $0 = \kappa_1 - \kappa_m \leq 1$ so that $\mathcal{W}_\circ^{1\times 1}$ is the set of $M \in \mathcal{W}$ such that $M(z) \neq 0$ for all $z \in \mathbb{T}$. Notice that for every element of $\mathcal{W}_\circ^{m\times m}$, the partial indices are either all nonnegative, all nonpositive, or all zero. This implies that for every element of $\mathcal{W}_\circ^{m\times m}$,

$$\operatorname{sign}(\kappa_i) = \operatorname{sign}\left(\sum_{i=1}^{m} \kappa_i\right), \qquad i = 1, \ldots, m,$$

where sign is equal to 1, −1, or 0 according to whether the argument is positive, negative, or zero, respectively. Since $\sum_{i=1}^{m} \kappa_i$ is the winding number of $\det(M(z))$ as z traverses the unit circle counterclockwise (Gohberg and Fel'dman, 1974, Thm. VIII.3.1(c)), we can easily determine the sign of the partial indices of elements of $\mathcal{W}_\circ^{m\times m}$ by looking at the winding number. The importance of this fact will become clear in the next section when we combine it with the following fact.

THEOREM 2. *If $\mathcal{W}^{m\times m}$ is endowed with the μ-essential supremum norm, then $\mathcal{W}_\circ^{m\times m}$ is open and dense in $\{M \in \mathcal{W}^{m\times m} : \det(M(z)) \neq 0$ for all $z \in \mathbb{T}\}$.*

Proof. See Theorems 1.20 and 1.21 of Gohberg, Kaashoek, and Spitkovsky (2003b). □

Theorem 2 implies that the generic or typical element of $\mathcal{W}^{m\times m}$ that admits a Wiener–Hopf factorization is an element of $\mathcal{W}_\circ^{m\times m}$. Said differently, the elements of $\mathcal{W}^{m\times m} \backslash \mathcal{W}_\circ^{m\times m}$ are nongeneric or exceptional in the space of Wiener–Hopf factorizable elements of $\mathcal{W}^{m\times m}$.

To see the role played by Wiener–Hopf factorization, consider system (6) again. In the first case, M factorized as

$$M_+(z) = 1 - \delta^{-1} z, \qquad M_0(z) = 1 \qquad M_-(z) = -a\delta(1 - \gamma z^{-1}).$$

In the second case, M factorized as

$$M_+(z) = 1, \qquad M_0(z) = z \qquad M_-(z) = a(1 - \delta z^{-1})(1 - \gamma z^{-1}).$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

In the third case, M factorized as

$$M_+(z) = a\delta\gamma(1-\delta^{-1}z)(1-\gamma^{-1}z), \qquad M_0(z) = z^{-1} \qquad M_-(z) = 1.$$

In the fourth case, $M(w) = 0$ for some $w \in \mathbb{T}$ and there does not exist a solution in general. Notice that the case of existence and uniqueness is associated with a partial index of zero, the case of existence and nonuniqueness is associated with a positive partial index, and the cases of nonexistence in general are associated with a negative partial index and/or a zero of $M(z)$. These associations are not accidental as we will see in the next section.

5. EXISTENCE AND UNIQUENESS

Given the Wiener–Hopf factorization techniques, we can now address the general LREM problem in the frequency domain. Our first task is to define existence and uniqueness of solutions to the LREM problem in the time domain.

DEFINITION 4. *An LREM is a pair* $(M,N) \in \mathcal{W}^{m\times m} \times \mathcal{W}^{m\times n}$, *expressed formally as*

$$\sum_{s=-\infty}^{\infty} M_s E_t X_{t+s} = \sum_{s=-\infty}^{\infty} N_s E_t \xi_{t+s}, \qquad t \in \mathbb{Z}, \tag{9}$$

where ξ *is exogenous and* X *is endogenous.*

Expression (9) is understood as a formal set of structural equations whose mathematical meaning we will postpone to Definition 5 below. The structural equations relate current, expected, and lagged values of the endogenous process X to current, expected, and lagged values of an exogenous process ξ. In particular, $E_t X_{t+s}$ is understood to mean the expectation of X_{t+s} at time t if $s > 0$ and $X_{t-|s|}$ if $s \le 0$; $E_t \xi_{t+s}$ is understood similarly. The notion of expectation pertinent to us will be made clear in Definition 5. A textbook example is the New Keynesian LREM,

$$\begin{aligned} X_{1t} &= \theta_1 E_t X_{1,t+1} + \theta_2 X_{2t}, \\ X_{2t} &= E_t X_{2,t+1} + \theta_3 X_{3t}, \\ X_{3t} &= \theta_4 X_{1,t-1} + \theta_5 X_{2,t-1} + \theta_6 \xi_{1t}, \end{aligned} \tag{10}$$

where X_1, X_2, and X_3 consist of inflation, output, and the interest rate, respectively; ξ is interpreted as a policy shock; and $\theta = (\theta_1, \ldots, \theta_6)$ is a parameter. The first equation of (10) is the New Keynesian Phillips curve, relating current inflation to one-period-ahead expected inflation and current output, and results from the inter-temporal optimization of firms; the second equation relates current output to one-period-ahead expected output and the interest rate and results from the inter-temporal optimization of households; finally, the third equation is the policy rule by which the monetary authority sets the interest rate (see, e.g., Galí, 2015). LREM (10) is represented by the pair,

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

$$M(z,\theta) = \begin{bmatrix} 1-\theta_1 z & -\theta_2 & 0 \\ 0 & 1-z & -\theta_3 \\ -\theta_4 z^{-1} & -\theta_5 z^{-1} & 1 \end{bmatrix}, \qquad N(z,\theta) = \begin{bmatrix} 0 \\ 0 \\ \theta_6 \end{bmatrix}.$$

Occasionally, researchers also consider solutions driven by sunspots, economic forces that do not appear explicitly in the system (their associated columns of N_s are equal to zero) but are nevertheless considered to influence the behavior of the solution; we discuss these solutions further in Section 6.

The class of models considered in Definition 4 includes the class of structural equation models (M and N are constant), the class of VARMA models (M and N are matrix polynomials in z^{-1}), and all of the LREMs considered in Canova (2011), DeJong and Dave (2011), and Herbst and Schorfheide (2016). For example, the model of Smets and Wouters (2007), studied in Section 6.2 of Herbst and Schorfheide (2016), consists of $m = 14$ endogenous variables and $n = 7$ exogenous shocks.

Now in order to endow (9) with mathematical meaning, we must first note that LREMs take as inputs not only realizations of exogenous inputs ξ but also a spectral measure F:

$$(\dots, \xi_{t-1}(\omega), \xi_t(\omega), \xi_{t+1}(\omega), \dots, F) \longrightarrow \boxed{(M,N)} \longrightarrow (\dots, X_{t-1}(\omega), X_t(\omega), X_{t+1}(\omega), \dots)$$

That is because the output X of an LREM solves a system of equations in past, present, as well as expected values of X. In the time-domain perspective on LREMs, the role of F is played by a filtration with respect to which conditional expectations can be computed (Al-Sadoon, 2018, Def. 4.2). Here, conditional expectations are substituted by linear projections, the more convenient analogue to expectations in the frequency domain and, in order to compute projections, F needs to be specified as well. We will see that the transfer function of solutions requires the triple (M,N,F), while the realizations of the output require, additionally, the realizations of the inputs. In following this system-theoretic approach to LREMs, therefore, we will often use phrases such as "for every spectral measure F" or "for every covariance stationary process ξ."

DEFINITION 5. *Let ξ be a zero-mean, n-dimensional, covariance stationary process with spectral measure F, and let (M,N) be an LREM. A solution to (M,N) (or equation (9)) is an m-dimensional covariance stationary process X, causal in ξ, and satisfying*

$$\sum_{s=-\infty}^{\infty} M_s \boldsymbol{P}(X_{t+s} | \mathscr{H}_t^m) = \sum_{s=-\infty}^{\infty} N_s \boldsymbol{P}(\xi_{t+s} | \mathscr{H}_t^n), \qquad t \in \mathbb{Z}, \tag{11}$$

where the series converge in $\mathscr{H}^m$. We say that (M,N) has no solution in general if it is possible to find a ξ such that no solution to (M,N) exists. A solution X is unique if whenever Y is also a solution, then $X_t = Y_t$ for all $t \in \mathbb{Z}$.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

By Definition 5, if a unique solution (or at least a unique selection from among all possible solutions) exists for every covariance stationary ξ, then the LREM is a mathematical system that transforms arbitrary covariance stationary inputs into covariance stationary outputs (Kalman, Arbib, and Falb, 1969; Kailath, 1980; Caines, 1988; Hannan and Deistler, 1988; Sontag, 1998). If no solution exists for a given ξ, then the LREM is said to have no solution in general (it will often be convenient to consider purely deterministic ξ to show that no solution exists in general). The restriction to covariance stationarity involves no loss of generality as X and ξ describe deviations away from steady state in modern LREMs (see, e.g., Galí, 2015; nonstationary LREMs are studied in Al-Sadoon, 2018).

Definition 5 also restricts attention to causal solutions. This is because the purpose of an LREM is to explain the behavior of economic variables in terms of past and present economic shocks and to obtain impulse responses. It makes little sense to consider models where current inflation is determined by a future shock to technology, for example. Note, however, that we do not impose invertibility of ξ in terms of X as it is not necessary for our purposes, although it does become necessary for estimation purposes (see Al-Sadoon and Zwiernik, 2019).

As in the previous section, it suffices to solve the $t = 0$ equation of (11),

$$\sum_{s=-\infty}^{\infty} M_s \boldsymbol{P}(X_s|\mathscr{H}_0^m) = \sum_{s=-\infty}^{\infty} N_s \boldsymbol{P}(\xi_s|\mathscr{H}_0^n).$$

For if X is a covariance stationary process causal in ξ and satisfies this system, applying the forward shift operator t times to each equation, we obtain (11). Of course, the forward shift operator commutes with the summation because the sum converges in $\mathscr{H}^m$. Let X_0 have the spectral characteristics $\phi \in H_0^m$. Then the frequency-domain equivalent is

$$\sum_{s=-\infty}^{\infty} M_s \boldsymbol{P}(z^s\phi|H_0^m) = \sum_{s=-\infty}^{\infty} N_s \boldsymbol{P}(z^s I_n|H_0^n). \tag{12}$$

Classical frequency-domain theory is built on the backward shift operator $\phi \mapsto z^{-1}\phi$. In order to solve for ϕ in (12), we will need an additional, closely related, operator.

Definition 6. *Define* $\boldsymbol{V}, \boldsymbol{V}^{(-1)} : H_0 \to H_0$ *to be the operators*

$$\boldsymbol{V} : \phi \mapsto \boldsymbol{P}(z\phi|H_0), \qquad \boldsymbol{V}^{(-1)} : \phi \mapsto z^{-1}\phi.$$

If $\kappa \geq 1$ *(resp.* $\kappa \leq -1$*), we denote by* $\boldsymbol{V}^\kappa$ *the operator* $\boldsymbol{V}$ *(resp.* $\boldsymbol{V}^{(-1)}$*) composed with itself* $|\kappa|$ *times and we define* $\boldsymbol{V}^0 = \boldsymbol{I}$*, the identity mapping.*

The following lemma (proof omitted) lists the most important properties of $\boldsymbol{V}$ and $\boldsymbol{V}^{(-1)}$.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

LEMMA 1. *The operators* $\boldsymbol{V}$ *and* $\boldsymbol{V}^{(-1)}$ *have the following properties:*

(i) $\boldsymbol{V}^* = \boldsymbol{V}^{(-1)}$.
(ii) $(\boldsymbol{V}\phi, \boldsymbol{V}\phi) \leq (\phi, \phi)$ *for all* $\phi \in H_0$.
(iii) $\left(\boldsymbol{V}^{(-1)}\phi, \boldsymbol{V}^{(-1)}\varphi\right) = (\phi, \varphi)$ *for all* $\phi, \varphi \in H_0$.
(iv) $\boldsymbol{V}^{(-1)}$ *is a right inverse of* $\boldsymbol{V}$.
(v) $\ker(\boldsymbol{V}) = H_0 \ominus z^{-1}H_0$.
(vi) $\dim \ker(\boldsymbol{V}) \leq n$.
(vii) *For all* $\kappa \in \mathbb{Z}$ *and* $\phi \in H_0$, $\boldsymbol{V}^\kappa \phi = \boldsymbol{P}(z^\kappa \phi | H_0)$.
(viii) *For* $\kappa > 1$, $\ker(\boldsymbol{V}^\kappa) = \ker(\boldsymbol{V}) \oplus \boldsymbol{V}^{(-1)} \ker(\boldsymbol{V}) \oplus \cdots \oplus \boldsymbol{V}^{1-\kappa} \ker(\boldsymbol{V})$.

Lemma 1(i) states that the backward shift operator $\boldsymbol{V}^{(-1)}$ is the adjoint of $\boldsymbol{V}$. Lemma 1(ii) implies that the operator norm of $\boldsymbol{V}$ is bounded above by 1. Lemma 1(iii) implies that $\boldsymbol{V}^{(-1)}$ is an isometry (Gohberg et al., 2003a, Thm. X.3.1). This implies that the operator norm of $\boldsymbol{V}^{(-1)}$ is equal to 1. Lemma 1(iv) establishes that $\boldsymbol{V}$ is right-invertible. Lemma 1(v) clarifies the obstruction to left-invertibility of $\boldsymbol{V}$ as $\ker(\boldsymbol{V})$ may be nontrivial. For example, when $F = \mu$ and $\phi \in \mathbb{C}$, then $\boldsymbol{V}^{(-1)}\boldsymbol{V}(\phi) = \boldsymbol{P}(\phi\,|z^{-1}, z^{-2}, \dots) = 0$. Since $H_0 \ominus z^{-1}H_0$ is the set of spectral characteristics associated with innovations to ξ, $\ker(\boldsymbol{V}) = \{0\}$ if and only if ξ is purely deterministic. Lemma 1(vi) expresses the intuitive fact that the dimension of the innovation space of a covariance stationary process is bounded above by the dimension of the process. Lemma 1(vii) is a convenient expression for iterates of the $\boldsymbol{V}$ operator. Finally, Lemma 1(viii) decomposes the kernel of $\boldsymbol{V}^\kappa$ into a direct sum generated by the kernel of $\boldsymbol{V}$. It follows, since $\boldsymbol{V}^{(-1)}$ is an isometry, that $\dim \ker(\boldsymbol{V}^\kappa) = \kappa \dim \ker(\boldsymbol{V})$. The time-domain analogue of the decomposition in Lemma 1(viii) is the familiar one from Broze, Gourieroux, and Szafarz (1985, 1995), where a process ν causal in ξ satisfies

$$\boldsymbol{P}(\nu_{t+\kappa} | \mathscr{H}_t^n) = 0, \qquad t \in \mathbb{Z},$$

if and only if

$$\nu_{t+\kappa} = \sum_{s=1}^{\kappa} \boldsymbol{P}(\nu_{t+\kappa} | \mathscr{H}_{t+s}^n) - \boldsymbol{P}(\nu_{t+\kappa} | \mathscr{H}_{t+s-1}^n), \qquad t \in \mathbb{Z}.$$

That is, if and only if $\nu_{t+\kappa}$ is representable as the sum of the prediction revisions between $t+1$ and $t+\kappa$ for all $t \in \mathbb{Z}$.

The fact that the operators $\boldsymbol{V}^s$ are uniformly bounded in the operator norm by 1 (Lemma 1(i) and (ii)) ensures that $\sum_{s=-\infty}^{\infty} M_s \boldsymbol{V}^s$ is a bounded linear operator on H_0 whenever $\sum_{s=-\infty}^{\infty} M_s z^s \in \mathcal{W}$ (Gohberg, Goldberg, and Kaashoek, 1990, Thm. I.3.2). More generally, we have the following definition, adopted from Gohberg and Fel'dman (1974).

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

DEFINITION 7. *For $M \in \mathcal{W}^{m\times n}$ with ijth element M_{ij}, define $\boldsymbol{M}_{ij} : H_0 \to H_0$ as*

$$\boldsymbol{M}_{ij} = \sum_{s=-\infty}^{\infty} M_{sij}\boldsymbol{V}^s, \qquad i = 1,\dots,m, \qquad j = 1,\dots,n,$$

where the series converges in the operator norm, and $\boldsymbol{M} : H_0^n \to H_0^m$ as

$$\boldsymbol{M}\phi = \begin{bmatrix} \boldsymbol{M}_{11} & \cdots & \boldsymbol{M}_{1n} \\ \vdots & & \vdots \\ \boldsymbol{M}_{m1} & \cdots & \boldsymbol{M}_{mn} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n \boldsymbol{P}(M_{1j}\phi_j|H_0) \\ \vdots \\ \sum_{j=1}^n \boldsymbol{P}(M_{mj}\phi_j|H_0) \end{bmatrix} = \boldsymbol{P}(M\phi|H_0^m).$$

For $M \in \mathcal{W}^{m\times n}$, $\|\boldsymbol{M}\phi\|_{H^m} = \|\boldsymbol{P}(M\phi|H_0^m)\|_{H^m} \le \|M\phi\|_{H^m} \le \|M\|_\infty\|\phi\|_{H^n}$, thus the operator norm of $\boldsymbol{M}$ is bounded by $\|M\|_\infty$. Note that by Lemma 1(i), the adjoint of $\boldsymbol{M}$ is

$$\boldsymbol{M}^* : \varphi \mapsto \boldsymbol{P}(M^*\varphi|H_0^n),$$

where $M^* \in \mathcal{W}^{n\times m}$ is given by $(M(z))^*$ for $z \in \mathbb{T}$.

Definition 7 allows us to express (12) more compactly as

$$\boldsymbol{M}\phi = \boldsymbol{N}I_n. \tag{13}$$

Recall that the spectral characteristic of ξ is I_n. Thus, we have arrived at a linear equation in the Hilbert space H_0^m. Equations (2), (4), and (6) are clearly special cases of (13). It is also instructive to consider the special case where (M,N) is a VARMA model, so $\boldsymbol{M}\phi = \boldsymbol{P}(M\phi|H_0^m) = M\phi$ and $\boldsymbol{N}I_n = \boldsymbol{P}(N|H_0^m) = N$, and the system reduces to $M\phi = N$, a problem that is very well understood (Hannan and Deistler, 1988, Sects. 1.1 and 1.2).

A more delicate analysis is required for the general case. Luckily Section 3 hints toward a solution. First, we obtain a Wiener–Hopf factorization,

$$M = M_+M_0M_-.$$

By Theorem 1, this factorization exists if $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$. Then $\boldsymbol{M}_+$, $\boldsymbol{M}_0$, and $\boldsymbol{M}_-$ can be defined as in Definition 7 and it is easily checked that

$$\boldsymbol{M} = \boldsymbol{M}_+\boldsymbol{M}_0\boldsymbol{M}_-,$$

a fact that at first seems trivial until one recalls that $\boldsymbol{V}^i\boldsymbol{V}^j$ is not generally equal to $\boldsymbol{V}^{i+j}$ when $i < 0 < j$ (Lemma 1(iv) and (v)). Then (13) can be broken up into the system of equations,

$$\boldsymbol{M}_+\psi = \boldsymbol{N}I_n, \qquad \boldsymbol{M}_0\varphi = \psi, \qquad \boldsymbol{M}_-\phi = \varphi. \tag{14}$$

Solving each of these systems is completely understood in the Wiener–Hopf factorization literature (Gohberg and Fel'dman, 1974; Clancey and Gohberg, 1981). Indeed, the first system of equations involves only the $\boldsymbol{V}$ operator and is solved as in the Cagan model, while the third system of equations involves only

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

the $\boldsymbol{V}^{(-1)}$ operator and is solved as in the autoregressive model. To see this, note that since $M_+^{-1}, M_-^{-1} \in \mathcal{W}^{m\times m}$, the operators

$$\boldsymbol{M}_+^{-1} : \phi \mapsto \boldsymbol{P}(M_+^{-1}\phi|H_0^m) \qquad\qquad \boldsymbol{M}_-^{-1} : \phi \mapsto \boldsymbol{P}(M_-^{-1}\phi|H_0^m)$$

are bounded and linear on H_0^m. By Lemma 1(vii), since $M_+, M_+^{-1} \in \mathcal{W}_+^{m\times m}$, we have that, for all $\phi \in H_0^m$,

$$\boldsymbol{M}_+\boldsymbol{M}_+^{-1}\phi = \boldsymbol{P}(M_+\boldsymbol{P}(M_+^{-1}\phi|H_0^m)|H_0^m) = \boldsymbol{P}(M_+M_+^{-1}\phi|H_0^m) = \boldsymbol{P}(\phi|H_0^m) = \phi,$$
$$\boldsymbol{M}_+^{-1}\boldsymbol{M}_+\phi = \boldsymbol{P}(M_+^{-1}\boldsymbol{P}(M_+\phi|H_0^m)|H_0^m) = \boldsymbol{P}(M_+^{-1}M_+\phi|H_0^m) = \boldsymbol{P}(\phi|H_0^m) = \phi.$$

On the other hand, since $M_-, M_-^{-1} \in \mathcal{W}_-^{m\times m}$, $\boldsymbol{M}_- : \phi \mapsto M_-\phi$ and $\boldsymbol{M}_-^{-1} : \phi \mapsto M_-^{-1}\phi$ for every $\phi \in H_0^m$ (i.e., they are multiplication operators). It follows that

$$\boldsymbol{M}_-\boldsymbol{M}_-^{-1}\phi = M_-M_-^{-1}\phi = \phi,$$
$$\boldsymbol{M}_-^{-1}\boldsymbol{M}_-\phi = M_-^{-1}M_-\phi = \phi.$$

We have established that

$$\boldsymbol{M}_+\boldsymbol{M}_+^{-1} = \boldsymbol{M}_+^{-1}\boldsymbol{M}_+ = \boldsymbol{M}_-\boldsymbol{M}_-^{-1} = \boldsymbol{M}_-^{-1}\boldsymbol{M}_- = \boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity mapping on H_0^m. The first and third equations of (14) are therefore uniquely solvable as

$$\psi = \boldsymbol{M}_+^{-1}\boldsymbol{N}I_n, \qquad\qquad \phi = \boldsymbol{M}_-^{-1}\varphi.$$

It remains to solve for φ in (14). Since $\boldsymbol{M}_0$ is a diagonal operator matrix with ith diagonal entry equal to $\boldsymbol{V}^{\kappa_i}$, existence and uniqueness of solutions to (13) depend on the partial indices of M. Notice that if all of the partial indices of M are equal to zero, then $\boldsymbol{M}_0 = \boldsymbol{I}$, and (13) has the unique solution $\phi = \boldsymbol{M}_-^{-1}\boldsymbol{M}_+^{-1}\boldsymbol{N}I_n$. If all the partial indices of M are nonnegative, the operators $\boldsymbol{V}^{\kappa_i}$ above are right invertible, which implies that $\boldsymbol{M}_0$ is right invertible; in that case, a right inverse of $\boldsymbol{M}_0$ is given as

$$\boldsymbol{M}_0^{(-1)} = \begin{bmatrix} \boldsymbol{V}^{(-\kappa_1)} & 0 & \cdots & 0 \\ 0 & \boldsymbol{V}^{(-\kappa_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{V}^{(-\kappa_m)} \end{bmatrix}. \tag{15}$$

If $\kappa_1 > 0$ and $\kappa_m \geq 0$, there are generally infinitely many other right inverses of $\boldsymbol{M}_0$ and therefore infinitely many right inverses of $\boldsymbol{M}$; indeed, every right inverse of $\boldsymbol{M}$ is of the form $\boldsymbol{M}_-^{-1}\boldsymbol{M}_0^{(-1)}\boldsymbol{M}_+^{-1}$ for some right inverse $\boldsymbol{M}_0^{(-1)}$ of $\boldsymbol{M}_0$.

The theoretical foundations for existence and uniqueness are now complete and all that remains is to apply well-known cookie-cutter results from functional analysis along with the simple techniques we employed in Section 3.

LEMMA 2. *If (M,N) is an LREM, $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, and M has a Wiener–Hopf factorization $M_+M_0M_-$, then for every spectral measure F, $\boldsymbol{M}$ is*

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

one-to-one if the partial indices of M are nonpositive and $\boldsymbol{M}$ *is onto if the partial indices of M are nonnegative. In the latter case, the general form of solutions to (13) is*

$$\phi = \boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \boldsymbol{M}_-^{-1}\psi, \tag{16}$$

where

$$\boldsymbol{M}^{(-1)} = \boldsymbol{M}_-^{-1}\boldsymbol{M}_0^{(-1)}\boldsymbol{M}_+^{-1}, \tag{17}$$

$\boldsymbol{M}_0^{(-1)}$ *is a right inverse of* $\boldsymbol{M}_0$, $\psi \in \ker(\boldsymbol{M}_0)$, *and* $\dim(\ker(\boldsymbol{M})) = \dim(\ker(\boldsymbol{V}))\sum_{i=1}^{m}\kappa_i$.

Proof. Proposition 2° of Section VIII.4 of Gohberg and Fel'dman (1974) applied to M^* together with Lemma 1 implies that

$$\dim(\operatorname{coker}(\boldsymbol{M})) = -\dim(\ker(\boldsymbol{V}))\sum_{\kappa_i<0}\kappa_i, \quad \dim(\ker(\boldsymbol{M})) = \dim(\ker(\boldsymbol{V}))\sum_{\kappa_i>0}\kappa_i.$$

Thus, $\boldsymbol{M}$ is one-to-one if $\kappa_1 \leq 0$ and onto if $\kappa_m \geq 0$. In the latter case, (16) is clearly a solution to (13). On the other hand, if ϕ is a solution to (13), define

$$\psi = \boldsymbol{M}_-\phi - \boldsymbol{M}_0^{(-1)}\boldsymbol{M}_+^{-1}\boldsymbol{N}I_n.$$

Clearly $\psi \in H_0^m$. Finally,

$$\boldsymbol{M}_0\psi = \boldsymbol{M}_0\boldsymbol{M}_-\phi - \boldsymbol{M}_+^{-1}\boldsymbol{N}I_n = \boldsymbol{M}_+^{-1}(\boldsymbol{M}\phi - \boldsymbol{N}I_n) = 0. \qquad \square$$

Basic linear algebra implies that there exists a solution to (13) if and only if $\boldsymbol{N}I_n$ is in the image of $\boldsymbol{M}$ and a solution is unique if and only if $\ker(\boldsymbol{M}) = \{0\}$. Thus, Lemma 2 provides a sufficient condition for existence and necessary and sufficient conditions for uniqueness of solutions to (13) irrespective of the exogenous inputs. Recall that we insist on conditions invariant with respect to F in keeping with our system-theoretic approach to LREMs. Of course, restricting attention to a particular F, we can say slightly more. For example, if the partial indices are nonnegative and F has the property that $\ker(\boldsymbol{V}) = \{0\}$, then by Lemma 1(viii), $\ker(\boldsymbol{M}_0) = \{0\}$ and there is a unique solution to (13). That is to say, there can be no multiplicity of solutions for a perfectly predictable ξ. This point is made in a different context in Al-Sadoon (2018, p. 641).

In the special case of a VARMA model, Lemma 2 reduces to the classical result that if $\det(M(z)) \neq 0$ for all $|z| \geq 1$, there exists a unique causal covariance stationary solution for every spectral measure F (Hannan and Deistler, 1988, Sects. 1.1 and 1.2). This is due to the fact that in that case $M_+ = M_0 = I_m$ and $M_- = M$ is a Wiener–Hopf factorization of M. Note that the problem of uniqueness does not arise in the case of VARMA.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Consider the case of existence and nonuniqueness ($\kappa_1 > 0$ and $\kappa_m \geq 0$) more closely. Lemma 1(v) and (viii) implies that $\ker(\boldsymbol{M}_0)$ is made up of spectral characteristics corresponding to arbitrary innovation processes (see Appendix A for a detailed analysis). Thus, Lemma 2 states that the set of all solutions to (13) is generated by spectral characteristics corresponding to arbitrary innovation processes. The dimension of $\ker(\boldsymbol{M})$ is the dimension of the innovation space of ξ multiplied by the winding number of $\det(M)$. This number is obtained in Funovits (2017, 2020) based on the Sims (2002) framework (see also Sorge, 2019).

Lemma 2 begs the question of what happens if zeros are present on the unit circle. This is addressed in the next result.

LEMMA 3. *If* (M,N) *is an LREM,* $\det(M(w)) = 0$ *for some* $w \in \mathbb{T}$, *and* $\operatorname{rank}[\, M(w) \quad N(w) \,] = m$, *there exists a spectral measure* F *such that (13) has no solution in* H_0^m.

Proof. Let $0 \neq x \in \mathbb{C}^{1\times m}$ satisfy $xM(w) = 0$ and choose $F = \delta_w I_n$. If a solution $\phi \in H_0^m$ to (13) exists, it must satisfy $\|\phi\|_{H^m}^2 = \sum_{j=1}^m \int \phi_j dF \phi_j^* = \|\phi(w)\|_{\mathbb{C}^{m\times n}}^2 < \infty$. Since $xM(z)\phi(z) = 0$ for F–a.e. $z \in \mathbb{T}$, it follows that $x\boldsymbol{M}\phi = \boldsymbol{P}(xM\phi|H_0) = 0$. This implies that $0 = x\boldsymbol{N}I_n = \boldsymbol{P}(xN|H_0)$. Since $xN(z) = xN(w)$ for F–a.e. $z \in \mathbb{T}$ and $xN(w) \in H_0$, $xN(w) = \boldsymbol{P}(xN|H_0) = 0$. This implies that $x[\, M(w) \quad N(w) \,] = 0$, a contradiction. □

The basic idea behind Lemma 3 is that when $\det(M(w)) = 0$ for some $w \in \mathbb{T}$, the system has a form of instability akin to that of a resonance frequency in a mechanical system. When such a system is subjected to an input oscillating at frequency $\arg(w)$, its output cannot be covariance stationary. In particular, a mechanical system will oscillate with increasing amplitude until failure (Arnold, 1973, p. 183). In the parlance of system theory, the system is said to have an unstable mode (Sontag, 1998, Chap. 5).

The rank condition on $[\, M(w) \quad N(w) \,]$ in Lemma 3 permits inputs at frequency $\arg(w)$ to excite the instability in the system. In the VARMA literature, it is typically assumed that $\operatorname{rank}[\, M(z) \quad N(z) \,] = m$ for all $z \in \mathbb{C}$ (Hannan and Deistler, 1988, Chap. 2). In the systems theory literature, similar conditions characterize controllability of the output in terms of the input (Kailath, 1980, Chap. 6). Without a condition of this sort, there may be no input that can excite the system's instability. For example, the LREM $(M,N) = (1 - z^{-1}, 1 - z^{-1})$ has a solution $\phi = 1$ for any F. Note that this condition permits oscillatory inputs to excite instability but not necessarily white noise inputs. For example, the LREM $(M,N) = (-z + 3 - 2z^{-1}, 3 - 2z^{-1})$ has $M(1) = 0$ and $N(1) = 1$ so that the conditions of Lemma 3 are satisfied, but the instability of the system cannot be excited by a white noise input because, as is easily checked, $\phi = 1$ is a solution to (13) when $F = \mu$. It is possible to formulate a different condition on N that will permit white noise inputs to excite instability in the system and, indeed, much more can be said about stability in LREMs. However, a general analysis of stability of LREMs is outside the scope of this paper and is left for future research.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

We are now ready to state the following result.

THEOREM 3 (Onatski's first theorem). *If (M,N) is an LREM, $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, and M has a Wiener–Hopf factorization $M_+M_0M_-$, then:*

(i) *If the partial indices of M are nonnegative, then for every covariance stationary process ξ, there exists a solution X to (M,N). The general form of solutions to (M,N) is*

$$X_t = \int z^t \boldsymbol{M}^{(-1)} \boldsymbol{N} I_n d\Phi + \int z^t \boldsymbol{M}_-^{-1} \psi d\Phi, \quad t \in \mathbb{Z},$$

where $\boldsymbol{M}^{(-1)}$ is a right inverse of $\boldsymbol{M}$, $\psi \in \ker(\boldsymbol{M}_0)$, and the dimension of the solution space is the dimension of the innovation space of ξ times the winding number of $\det(M)$.

(ii) *If the partial indices of M are all equal to zero, there is a unique solution for every covariance stationary process ξ given by*

$$X_t = \int z^t \boldsymbol{M}_-^{-1} \boldsymbol{M}_+^{-1} \boldsymbol{N} I_n d\Phi, \quad t \in \mathbb{Z}.$$

(iii) *If M has a negative partial index and the zeroth Fourier coefficient of the last row of $[M_+^{-1}N]_-$ is nonzero, then there exists no solution to (M,N) in general.*

Proof. (i) and (ii) follow immediately from Lemma 2 and the spectral representation theorem. (iii) We claim that for $F = \mu I_n$, there exists no solution to (M,N). For any solution ϕ must satisfy $\boldsymbol{M}_0\boldsymbol{M}_-\phi = \boldsymbol{M}_+^{-1}\boldsymbol{N}I_n = [M_+N]_-$ as $\{z^s e_j : j = 1,\dots,n, s \in \mathbb{Z}\}$ is orthonormal for our choice of F. But the last row of this equation is $z^{\kappa_m} e_m M_- \phi = e_m [M_+^{-1}N]_-$, which evidently cannot hold if $\kappa_m < 0$, as the zeroth Fourier coefficient of the left-hand side is zero, while that of the right-hand side is not by assumption. □

The relationship between partial indices and existence and uniqueness of solutions to LREMs was first discovered by Onatski (2006). Theorem 3 generalizes Onatski (2006) by allowing for arbitrary covariance stationary input and nonconstant N. The general expression of solutions in (i) is similar to the one obtained in Theorem 4.1(ii) of Al-Sadoon (2018); Lubik and Schorfheide (2003), Farmer et al. (2015), Bianchi and Nicolò (2021), and Al-Sadoon (2020) provide algorithms for computing these solutions.

Note that the analogue of (iii) in Onatski (2006) is incorrect: if N (Γ in Onatski's notation) is constant and nonzero, and M (A in Onatski's notation) has a negative (positive in Onatski's framework) partial index, it does not follow that a solution fails to exist. A counterexample is $(M,N) = \left(\begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$, which has the unique solution $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for any choice of F. It is easy to check, however, that Onatski's statement is correct in the scalar case $m = 1$. Our additional condition on $[M_+^{-1}N]_-$, like the coprimeness condition in Lemma 3, permits us to use white noise as

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

an input to the system for which there cannot be an admissible output (different conditions can also be given). We will see shortly that this condition is generic.

The unit-root case can be stated as the following.

THEOREM 4. *If* (M,N) *is an LREM,* $\det(M(w)) = 0$ *for some* $w \in \mathbb{T}$*, and* $\operatorname{rank}[\, M(w) \quad N(w) \,] = m$*, then there exists no solution to* (M,N) *in general.*

Proof. Follows from Lemma 3 and the spectral representation theorem. □

We can also state the following result for generic systems.

THEOREM 5 (Onatski's second theorem). *For a generic LREM* (M,N) *with* $\det(M(z)) \neq 0$*, for all* $z \in \mathbb{T}$*, there exist possibly infinitely many solutions, a unique solution, or no solution in general according to whether* $\det(M(z))$ *winds around the origin a positive, zero, or negative number of times as* z *traverses* $\mathbb{T}$ *counterclockwise.*

Proof. By Theorem 2, $\mathcal{W}_{\circ}^{m\times m}$ is generic in the space of Wiener–Hopf factorizable elements of $\mathcal{W}^{m\times m}$. For fixed $M \in \mathcal{W}_{\circ}^{m\times m}$ with Wiener–Hopf factorization $M = M_+M_0M_-$, the generic element of $\mathcal{W}^{m\times n}$ has a zeroth Fourier coefficient of $e_m[M_+^{-1}N]_-$ that is nonzero (i.e., $\int e_m[M_+^{-1}N]_-d\mu \neq 0$). Indeed, when $\mathcal{W}^{m\times n}$ is endowed with the μ-essential supremum norm, $N \mapsto \int e_m[M_+^{-1}N]_-d\mu$ is a continuous mapping from $\mathcal{W}^{m\times n}$ to $\mathbb{C}^{1\times n}$. The result then follows from Theorem 3(iii). □

In closing, it is important to note that the assumptions underlying existence and uniqueness in this paper are weaker than in all of the previous literature on frequency-domain solutions of LREMs, namely the work of Whiteman (1983), Onatski (2006), Tan and Walker (2015), Tan (2019), and Meyer-Gohde and Tzaawa-Krenzler (2023). These works require the exogenous process to have a purely nondeterministic Wold (1938) representation, an assumption that is demonstrated to be unnecessary. The weaker assumptions of this paper have also facilitated discussion of zeros of M, an aspect of the theory absent from the previous literature. The advantage of these stronger assumptions, however, is that they do permit closed-form expressions of solutions. See Appendix B for a more detailed discussion.

6. ILL-POSEDNESS AND REGULARIZATION

We have seen in the previous section that LREMs may have infinitely many solutions. In order to estimate and conduct inference, various selection mechanisms have been proposed in the macroeconometrics literature to associate a unique solution with any given set of parameters. We will review these proposals and show that, unfortunately, continuity of the selected solutions is not guaranteed. This

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

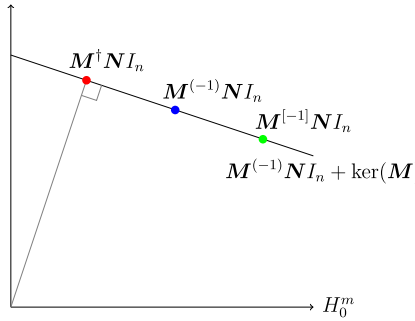


FIGURE 1. Solutions to linear rational expectations models.

leads to the development of a new regularized solution with guaranteed regularity properties.

6.1. Nonuniqueness

The first approach to nonuniqueness, proposed by Taylor (1977), selects the solution that minimizes the variance of the price variable in the model if one exists. Taylor motivates this solution by arguing that collective rationality of economic agents will naturally lead them to coordinate their activities to achieve this equilibrium. Unfortunately, this provides no guidance for models in which indeterminacy afflicts non-price variables.

The second approach to nonuniqueness commits to a particular algorithm for obtaining a solution and ignores all other solutions. That is, it commits to a particular choice of right inverse to $\boldsymbol{M}$, say $\boldsymbol{M}^{(-1)}$, and selects the solution $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n$, ignoring all other solutions $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \ker(\boldsymbol{M})$. This is the "minimum state variable" approach of McCallum (1983). Although this may appear to solve the problem of selecting a unique solution, there are generally infinitely many algorithms for obtaining solutions as there are infinitely many right inverses of $\boldsymbol{M}$. Thus, the concept of minimum state variable solution is not well defined without specifying the particular algorithm or right inverse to be used for the solution. Geometrically, there is a minimum state variable solution at every single point of $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \ker(\boldsymbol{M})$ (see Figure 1). Without sound economic reasoning for the choice of algorithm or right inverse, it cannot be argued that one minimum state variable solution obtained by one algorithm is a more appropriate choice than another, obtained by a different algorithm.

Finally, the modern approach to nonuniqueness, as exemplified by Lubik and Schorfheide (2003), Farmer et al. (2015), and Bianchi and Nicolò (2021), parametrizes the full set of solutions. Like the minimum state variable approach, it commits to a particular algorithm or right inverse of $\boldsymbol{M}$, say $\boldsymbol{M}^{(-1)}$, and represents each solution as a sum $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \chi$. The first part, $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n$, is interpreted

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

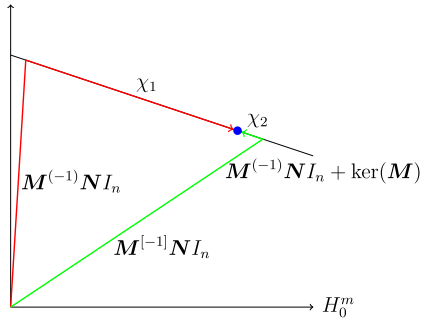


FIGURE 2. Indeterminacy of sunspots in solutions to linear rational expectations models.

as generated by fundamental economic forces (the ones that appear explicitly in the model (9)), while the second part, $\chi \in \ker(\boldsymbol{M})$, generated as we have seen by arbitrary innovation processes, is interpreted as non-fundamental sunspots (Farmer, 1999). The coordinates on $\ker(\boldsymbol{M})$ are appended to the parameters of (M,N) and both estimated by either frequentist or Bayesian methods. The geometry of the spectral approach allows us to see very clearly a serious conceptual problem with this approach. Various papers in the literature claim to present evidence of the importance of sunspots as drivers of macroeconomic activity by showing that the contribution of sunspots, as measured by the size of the estimated χ, is not insignificant empirically. However, it is clear from Figure 2 that the size of χ very much depends on the choice of algorithm or right inverse of $\boldsymbol{M}$: by one representation, $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \chi_1$, sunspots play a large role, and by another representation, $\boldsymbol{M}^{[-1]}\boldsymbol{N}I_n + \chi_2$, sunspots play a small role. Without a sound economic reason for the choice of algorithm or right inverse of $\boldsymbol{M}$, it is unclear that the contribution of sunspots is being measured correctly. Thus, the modern approach suffers from a similar difficulty as the minimum state variable approach.

6.2. Discontinuity

We have established in Lemma 2 that the set of all solutions to (13) is of the form $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \ker(\boldsymbol{M})$ for some right inverse $\boldsymbol{M}^{(-1)}$ of $\boldsymbol{M}$. In the course of proving Theorem 7 below, we will see that small variations of M in the μ-essential supremum norm lead to small variations in the orthogonal projection onto $\ker(\boldsymbol{M})$ in the operator norm. Unfortunately, small changes in M in the μ-essential supremum norm are not guaranteed to lead to small changes in $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n$ in the H^m-norm. Discontinuity can occur when M falls inside the nongeneric set $\mathcal{W}^{m\times m}\backslash\mathcal{W}_\circ^{m\times m}$. We will illustrate this discontinuity with the simplest possible example. It is, of course, possible to illustrate discontinuity using a more realistic example similar to (10), but doing so comes at the cost of analytic tractability, as solving even the simplest multivariate LREMs can be quite cumbersome.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

It is important to keep in mind that LREMs are typically parametrized purely on theoretical considerations, without any regard to statistical properties, so that elements of $\mathcal{W}^{m\times m}\backslash\mathcal{W}_{\circ}^{m\times m}$ are not expressly excluded from the parametrization (Onatski, 2006; Sims, 2007).

Consider the following example:

$$F=\mu I_2,\qquad M(z,\theta)=\begin{bmatrix} z^2 & 0\\ \theta z & 1\end{bmatrix},\qquad N(z)=\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}, \tag{18}$$

with $\theta\in\mathbb{R}$. Thus, ξ is a standardized white noise process.

In order to compute solutions, we will need to obtain a Wiener–Hopf factorization,

$$M(z,\theta)=\begin{cases}\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix} z^2 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}, & \theta=0,\\ \begin{bmatrix} 1 & z\\ 0 & \theta\end{bmatrix}\begin{bmatrix} z & 0\\ 0 & z\end{bmatrix}\begin{bmatrix} 0 & -\theta^{-1}\\ 1 & \theta^{-1}z^{-1}\end{bmatrix}, & \theta\neq 0.\end{cases}$$

Thus, there exists a (nonunique) solution for every $\theta\in\mathbb{R}$. Define $\boldsymbol{M}(\theta)$ as in Definition 7, $\phi\mapsto \boldsymbol{P}(M(\cdot,\theta)\phi|H_0^2)$. Define $\boldsymbol{M}(\theta)^{(-1)}$ as in (17) with $\boldsymbol{M}_0(\theta)^{(-1)}$ defined as in (15). We will restrict attention to minimum state variable solutions (i.e., $\psi=0$ in (16)),

$$\phi(\theta)=\boldsymbol{M}(\theta)^{(-1)}\boldsymbol{N}I_n=\begin{cases}\begin{bmatrix} z^{-2} & 0\\ 0 & 1\end{bmatrix}, & \theta=0,\\ \begin{bmatrix} z^{-2} & \theta^{-1}z^{-1}\\ -\theta z^{-1} & 0\end{bmatrix}, & \theta\neq 0.\end{cases}$$

This solution is clearly discontinuous at $\theta=0$. In particular,

$$\|\phi(\theta)-\phi(0)\|_{H^2}^2=\int|\theta^{-1}z^{-1}|^2d\mu+\int|\theta z^{-1}|^2d\mu+\int|1|^2d\mu=\theta^{-2}+\theta^2+1$$

tends to infinity as $\theta\to 0$. Thus, discontinuities can arise in the process of solving an LREM.

While this discontinuity may be quite jarring to readers acquainted with the LREM literature, from the Wiener–Hopf factorization literature point of view, there is nothing surprising about this discontinuity. Indeed, (18) is a minor modification of the example given in Section 1.5 of Gohberg et al. (2003b). The fact that only multivariate systems with nonunique solutions can exhibit this discontinuity may explain why discontinuity has not received sufficient attention in the LREM literature, as solving such models is difficult to do analytically. The heart of the problem is that, when $m>1$, there are points in $\mathcal{W}^{m\times m}$ at which partial indices are discontinuous, these are precisely the nongeneric set $\mathcal{W}^{m\times m}\backslash\mathcal{W}_{\circ}^{m\times m}$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

(Gohberg et al., 2003b, Cor. 1.22). That is, if $M \in \mathcal{W}^{m\times m}$ has partial indices satisfying $\kappa_1 - \kappa_m > 1$, then there are small (in the μ-essential supremum norm) changes in M that lead to a jump in M_0 and since M has undergone only a small change, $M_\pm$ must also jump. If we then choose a fixed right inverse $\boldsymbol{M}_0^{(-1)}$ in the solution (16) (e.g., choosing $\boldsymbol{M}_0^{(-1)}$ as in (15)), then the discontinuity can affect the solution. This is, unfortunately, what is done in all existing solution algorithms (Al-Sadoon, 2018, 2020). In our example, the partial indices of $M(z,\theta)$ are $(2,0)$ for $\theta = 0$ and $(1,1)$ for $\theta \neq 0$; $M(z,0) \in \mathcal{W}^{2\times 2}\backslash\mathcal{W}_\circ^{2\times 2}$ and $M(z,\theta) \in \mathcal{W}_\circ^{2\times 2}$ for $\theta \neq 0$. The discontinuity in the partial indices is what generates the discontinuity of the Wiener–Hopf factors, which in turn generates a discontinuity in the solution.

It bears emphasizing that this discontinuity is a feature of the mathematical problem, it is not a feature of the particular algorithm used to solve the problem; Al-Sadoon (2020) shows how it arises in the Sims (2002) framework, Al-Sadoon (2018) shows how it arises in a linear systems framework, and Appendix C shows how it arises when solving the system by hand. It is also important to note that the discontinuity of Wiener–Hopf factorization implies that there can exist no numerically stable way to compute it generally (see Example C.2 of the online supplement to Al-Sadoon (2018) for an illustration). While the elements of $\mathcal{W}_\circ^{m\times m}$ can be factorized using finite precision arithmetic, the elements of $\mathcal{W}^{m\times m}\backslash\mathcal{W}_\circ^{m\times m}$ cannot be factorized without infinite precision. Thus, it is a nonstarter to verify numerically whether a given system is generic or not in the process of estimation and inference. Note that an analogous problem arises in the computation of the Jordan canonical form, which is discontinuous at a nongeneric set of matrices in $\mathbb{C}^{m\times m}$ when $m > 1$ (Horn and Johnson, 1985, pp. 127–128); there, the recommendation is to compute the Schur canonical form instead; and in the next section, we will propose, similarly, to compute different solutions to LREMs than the ones proposed in the literature.

6.3. Regularization

Having established that the LREM problem in macroeconometrics is ill-posed, we now consider how to obtain economically meaningful solutions amenable to mainstream econometric techniques.

Perhaps the most natural solution to the ill-posedness problem is to avoid it altogether and restrict attention to systems with unique solutions (i.e., LREMs (M,N), where the partial indices of M are equal to zero). Al-Sadoon and Zwiernik (2019) have shown that unique solutions are not only continuous in the parameters of an LREM but also analytic (see the proof of Theorem 6.2). A less restrictive solution is to allow for nonuniqueness but restrict attention to generic systems (i.e., $\mathcal{W}_\circ^{m\times m}$). However, genericity cannot be taken for granted as both Onatski (2006) and Sims (2007) have warned and some models may be parametrized to always fall inside $\mathcal{W}^{m\times m}\backslash\mathcal{W}_\circ^{m\times m}$ (interestingly, Sims, 2007 is widely but erroneously considered to be a critique of Onatski, 2006; see Al-Sadoon, 2019). Moreover, we need differentiability, not just continuity, in order to ensure asymptotic normality

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

of extremum estimators (Hannan and Deistler, 1988; Pötscher and Prucha, 1997) as it facilitates the construction of confidence intervals as well as hypothesis testing (Newey and McFadden, 1994). Therefore, we opt for a more straightforward solution, regularization.

With the geometry of the spectral approach in view, one is led inexorably to consider the Tykhonov-regularized solution to (13), which minimizes the total variance, $\|X_0\|^2_{\mathscr{H}^m} = \|\phi\|^2_{H^m}$ among all solutions to (11),

$$\phi_{\boldsymbol{I}} = \arg\min\left\{\|\phi\|^2_{H^m} : \boldsymbol{M}\phi = \boldsymbol{N}I_n\right\}.$$

It can be shown that $\phi_{\boldsymbol{I}} = \boldsymbol{M}^{\dagger}\boldsymbol{N}I_n$, where $\boldsymbol{M}^{\dagger}$ is the Moore–Penrose inverse of $\boldsymbol{M}$ (Groetsch, 1977, p. 41). This takes a particularly simple form in our context.

LEMMA 4. *If $M \in \mathcal{W}^{m\times m}$, $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, and the partial indices of M are nonnegative, then*

$$\boldsymbol{M}^{\dagger} = \boldsymbol{M}^*(\boldsymbol{M}\boldsymbol{M}^*)^{-1}.$$

Proof. By Lemma 2, $\boldsymbol{M}$ is onto. It follows that $\boldsymbol{M}^{\dagger} = \boldsymbol{M}^*(\boldsymbol{M}\boldsymbol{M}^*)^{\dagger}$ (Groetsch, 1977, Thm. 2.1.5). Since $\boldsymbol{M}$ is onto, $\boldsymbol{M}^*$ is one-to-one and so $(\boldsymbol{M}\boldsymbol{M}^*)^{\dagger} = (\boldsymbol{M}\boldsymbol{M}^*)^{-1}$. □

Lemma 4 implies a very simple technique for computing the Tykhonov-regularized solution. Whenever a solution to (13) exists, the Tykhonov-regularized solution is obtained from the unique solution to the auxiliary block triangular frequency-domain system,

$$\begin{bmatrix} \boldsymbol{M}\boldsymbol{M}^* & 0 \\ -\boldsymbol{M}^* & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \varphi \\ \phi_{\boldsymbol{I}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{N}I_n \\ 0 \end{bmatrix}.$$

This implies that whenever a solution exists, the regularized solution is obtained uniquely by solving an auxiliary LREM. For example, in the mixed model from Section 3, the Tykhonov-regularized solution is the solution X to the auxiliary LREM,

$$\begin{aligned} a\bar{c}E_tY_{t+2} + (a\bar{b}+b\bar{c})E_tY_{t+1} + (|a|^2+|b|^2)Y_t + |c|^2E_{t-1}Y_t \\ + (b\bar{a}+c\bar{b})Y_{t-1} + c\bar{a}Y_{t-2} = \xi_t, \\ X_t = \bar{a}Y_{t-1} + \bar{b}Y_t + \bar{c}E_tY_{t+1}. \end{aligned}$$

This regularized solution exists and is unique whenever the original system satisfies the conditions for existence.

Note that when the partial indices of M are all equal to zero, $\boldsymbol{M}^{\dagger} = \boldsymbol{M}^*(\boldsymbol{M}\boldsymbol{M}^*)^{-1} = \boldsymbol{M}^*(\boldsymbol{M}^*)^{-1}(\boldsymbol{M})^{-1} = (\boldsymbol{M})^{-1}$. In other words, Tykhonov-regularization has no effect if the solution is unique.

More generally, we may consider

$$\phi_{\boldsymbol{L}} \in \arg\min\left\{\|\boldsymbol{L}\phi\|^2_{H^l} : \boldsymbol{M}\phi = \boldsymbol{N}I_n\right\},$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

where $\boldsymbol{L} : H^m \to H^l$ is a bounded linear operator chosen by the researcher. Next, we consider economic motivation for a number of choices of $\boldsymbol{L}$.

If the researcher wishes to shrink the variance of the ith component of X, they can set

$$\boldsymbol{L} : \phi \mapsto \phi_i.$$

Note that when the ith component is the price variable, we obtain the Taylor (1977) solution.

One can also shrink expected values of X. For example, if certain solutions obtained by the methods reviewed above yield expectations of output that are too variable relative to what one expects empirically, then one can impose this prior by using

$$\boldsymbol{L} : \phi \mapsto \boldsymbol{V}\phi_j,$$

where j is the coordinate corresponding to output in X.

Linear combinations of lagged, current, and expected values of coordinates of X can also be shrunk similarly. The operator

$$\boldsymbol{L} : \phi \mapsto \begin{bmatrix} (\boldsymbol{V} - 2\boldsymbol{I} + \boldsymbol{V}^{(-1)})\phi_1 \\ \vdots \\ (\boldsymbol{V} - 2\boldsymbol{I} + \boldsymbol{V}^{(-1)})\phi_m \end{bmatrix}$$

shrinks the second difference of all variables, imposing smoothness on solutions, similar to the idea of the Hodrick and Prescott (1997) filter.

Note that the operators above belong to the class of operators defined in Definition 7. Thus, we can more generally use any $L \in \mathcal{W}^{l\times m}$ to construct the weight

$$\boldsymbol{L} : \phi \mapsto \begin{bmatrix} \sum_{j=1}^m \boldsymbol{L}_{1j}\phi_j \\ \vdots \\ \sum_{j=1}^m \boldsymbol{L}_{lj}\phi_j \end{bmatrix}.$$

More importantly, regularization can allow the researcher to shrink not just across time but across frequencies. For example, the researcher may wish to shrink the spectrum of the solution toward frequencies of between $2\pi/32$ and $2\pi/4$ corresponding to business cycle fluctuations of period 4–32 quarters in quarterly data, that is, using

$$\boldsymbol{L} : \phi \mapsto L\phi, \qquad L(z) = \begin{cases} 0, & \pi/16 \le |\arg(z)| \le \pi/2, \\ I_m, & \text{otherwise.} \end{cases}$$

Finally, we can consider solutions that minimize a finite weighted sum of individual criteria as reviewed above,

$$a_1\|\boldsymbol{L}_1\phi\|^2_{H^{l_1}} + \cdots + a_d\|\boldsymbol{L}_d\phi\|^2_{H^{l_d}},$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

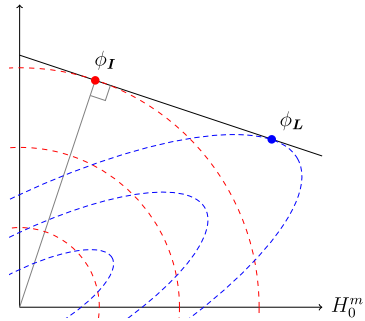


FIGURE 3. Regularized solutions to linear rational expectations models.

where $\boldsymbol{L}_i : H^m \to H^{l_i}$ and $a_i > 0$ for $i = 1, \ldots, d$. This would allow the researcher to impose d different criteria according to the weights $a_1, \ldots, a_d$. This can be achieved by using

$$\boldsymbol{L} : \phi \mapsto \begin{bmatrix} \sqrt{a_1}\boldsymbol{L}_1 \\ \vdots \\ \sqrt{a_d}\boldsymbol{L}_d \end{bmatrix} \phi.$$

The argument for regularization is the same as employed throughout the inverse and ill-posed problems literatures: if theory is insufficient to pin down a unique continuous solution, other information can be employed. In our case, regularization allows economically meaningful shrinkage criteria to select the best among all possible solutions or, what amounts to the same, it allows the researcher's priors about economic behavior to select the most appropriate solution. Regularization resolves the problem of selecting an economically grounded solution and, as we will soon see, also ameliorates the discontinuity problem.

Existence of regularized solutions is guaranteed whenever the given LREM satisfies the already minimal conditions for existence of solutions. That is because, according to Lemma 2, the set of all solutions is finite-dimensional and so the problem of minimizing $\|\boldsymbol{L}\phi\|^2_{H^l}$ subject to $\boldsymbol{M}\phi = \boldsymbol{N}I_n$ reduces to a problem of minimizing a nonnegative quadratic form (see Figure 3). Uniqueness of regularized solutions, on the other hand, is the subject of the next result.

LEMMA 5. *If (M, N) is an LREM, $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, the partial indices of M are nonnegative, and $\boldsymbol{L} : H^m \to H^l$ is a bounded linear operator, then*

$$\phi_{\boldsymbol{L}} = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L})\boldsymbol{M}^{\dagger}\boldsymbol{N}I_n,$$

where $\boldsymbol{L}|_{\ker(\boldsymbol{M})}$ is the restriction of $\boldsymbol{L}$ to $\ker(\boldsymbol{M})$, is a regularized solution to (13). $\phi_{\boldsymbol{L}}$ is the only element of $\arg\min\left\{\|\boldsymbol{L}\phi\|^2_{H^l} : \boldsymbol{M}\phi = \boldsymbol{N}I_n\right\}$ in H_0^m if and only if $\ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M}) = \{0\}$.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Proof. See Appendix D. □

In the course of proving Lemma 5, we find that the set of regularized solutions is $\phi_{\boldsymbol{L}} + \ker(\boldsymbol{M}) \cap \ker(\boldsymbol{L})$. Thus, regularization produces a unique solution if and only if $\ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M}) = \{0\}$. Geometrically, this condition requires the operator $\boldsymbol{L}$ to put weight on all directions of indeterminacy of the LREM; if $\ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M}) \neq \{0\}$, it will be possible to perturb $\phi_{\boldsymbol{L}}$ in any direction in $\ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M})$ to arrive at another regularized solution and uniqueness will fail. A host of other representations of $\phi_{\boldsymbol{L}}$ are obtained in Appendix D.

Some special cases of Lemma 5 are particularly instructive. When $\ker(\boldsymbol{L}) = \{0\}$, the regularized solution is unique regardless of $\boldsymbol{M}$. That is, when $\boldsymbol{L}$ puts weight on all directions in the solution space, the regularized solution is unique. An example of this is $\boldsymbol{L} = \boldsymbol{I}$, which produces the Tykhonov-regularized solution $\phi_{\boldsymbol{I}} = \boldsymbol{M}^{\dagger}\boldsymbol{N}\boldsymbol{I}_n$. When $\ker(\boldsymbol{M}) = \{0\}$ (i.e., the solution to the LREM is unique), the regularized solution is the unique solution regardless of $\boldsymbol{L}$. This is due to the fact that $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}$ is the zero operator (because $\boldsymbol{L}|_{\ker(\boldsymbol{M})}$ is the zero operator) and $\boldsymbol{M}^{\dagger} = (\boldsymbol{M})^{-1}$ so that $\phi_{\boldsymbol{L}} = (\boldsymbol{M})^{-1}\boldsymbol{N}\boldsymbol{I}_n$, the unique solution.

Recalling that $\phi_{\boldsymbol{L}}$ is a solution to the LREM, note how the mapping from parameters to the set of all solutions, $(M,N) \mapsto \phi_{\boldsymbol{L}} + \ker(\boldsymbol{M})$, compares to the mapping from parameters to the set of regularized solutions $(M,N) \mapsto \phi_{\boldsymbol{L}} + \ker(\boldsymbol{M}) \cap \ker(\boldsymbol{L})$. The difference is simply a matter of dimension reduction. It is important to note that neither mapping is generally one-to-one and solution sets associated with different parameters may intersect.

THEOREM 6. *If (M,N) is an LREM, $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, the partial indices of M are nonnegative, and $\boldsymbol{L} : H^m \to H^l$ is a bounded linear operator, then for every covariance stationary process ξ, there exists a solution X minimizing $\|\boldsymbol{L}\phi\|_{H^l}$ given by*

$$X_t = \int z^t (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L})\boldsymbol{M}^{\dagger}\boldsymbol{N}\boldsymbol{I}_n d\Phi, \quad t \in \mathbb{Z}.$$

The solution is unique if and only if $\ker(\boldsymbol{M}) \cap \ker(\boldsymbol{L}) = \{0\}$.

Proof. Follows from Lemma 5 and the spectral representation theorem. □

The expression for regularized solutions in Theorem 6 is primarily of theoretical interest. It will allow us to study continuity and differentiability with respect to underlying parameters. For estimation and inference, on the other hand, Al-Sadoon (2020) provides a numerical algorithm for computing regularized solutions in the Sims (2002) framework, which leads to equivalent algebraic rather than geometric criteria for existence and uniqueness.

We have shown that, with an appropriate choice of $\boldsymbol{L}$, the regularized solution can overcome the nonuniqueness problem. Our next result shows that regularized solutions also overcome the discontinuity problem. The continuity guarantee that we require for mainstream econometric methodology is not with respect to the H^m-

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

norm but with respect to the μ-essential supremum norm (see, e.g., the continuity results for bounded spectral densities and the Gaussian likelihood functions in Hannan, 1973; Deistler and Pötscher, 1984; Anderson, 1985). If we parametrize M and N as $M(z,\theta)$ and $N(z,\theta)$, then it is clear that we need $M(z,\theta)$ and $N(z,\theta)$ to be jointly continuous in z and θ. However, Green and Anderson (1987) note that this is not sufficient to ensure continuity of the Wiener–Hopf factors in the μ-essential supremum norm. Thus, we use Green and Anderson's idea of imposing control over $\frac{d}{dz}M(z,\theta)$ and $\frac{d}{dz}N(z,\theta)$.

THEOREM 7. *Let $M : \mathbb{T} \times \Theta \to \mathbb{C}^{m\times m}$ and $N : \mathbb{T} \times \Theta \to \mathbb{C}^{m\times n}$. Under the conditions:*

(i) $F = \mu I_n$.
(ii) $\Theta \subset \mathbb{R}^d$ *is an open set.*
(iii) $M(\cdot,\theta)$ *and* $N(\cdot,\theta)$ *are analytic in a neighborhood of* $\mathbb{T}$ *for every* $\theta \in \Theta$.
(iv) $M(z,\theta)$, $N(z,\theta)$, $\frac{d}{dz}M(z,\theta)$, *and* $\frac{d}{dz}N(z,\theta)$ *are jointly continuous at every* $(z,\theta) \in \mathbb{T} \times \Theta$.
(v) $\ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M}(\theta)) = \{0\}$ *for all* $\theta \in \Theta$.
(vi) $\det(M(z,\theta)) \neq 0$ *for all* $z \in \mathbb{T}$, *and the partial indices of* $M(z,\theta)$ *are all nonnegative for all* $\theta \in \Theta$.

Then $\phi_{\boldsymbol{L}}(\theta) = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M}(\theta))})^{\dagger}\boldsymbol{L})\boldsymbol{M}(\theta)^{\dagger}\boldsymbol{N}(\theta)I_n$ *is continuous in the* μ*-essential supremum norm.*

Proof. See Appendix D. □

The assumptions of Theorem 7 are quite strong relative to the discussion so far. However, the relevant case for most macroeconometric applications is the case where $F = \mu I_n$, while $M(z,\theta)$ and $N(z,\theta)$ are Laurent matrix polynomials of uniformly bounded degree. In this case, the continuity of the coefficients of $M(z,\theta)$ and $N(z,\theta)$ is sufficient to ensure conditions (iii) and (iv) of Theorem 7. The New-Keynesian model (10) satisfies these conditions, as do all of the LREMs in Canova (2011), DeJong and Dave (2011), or Herbst and Schorfheide (2016).

For the purpose of establishing asymptotic normality, we typically need not just continuity in the essential supremum norm but also differentiability in the essential supremum norm (i.e., the finite differential in θ converges to the infinitesimal differential in the essential supremum norm over $z \in \mathbb{T}$). This stronger form of differentiability allows us to differentiate under integrals, which appear in the asymptotics of maximum likelihood and generalized method of moments estimators. The following result provides exactly what we need.

THEOREM 8. *Let p be a positive integer. Under assumptions (i)–(vi) of Theorem 7,*

(vii) $M(z,\theta)$, $N(z,\theta)$, $\frac{d}{dz}M(z,\theta)$, *and* $\frac{d}{dz}N(z,\theta)$ *are jointly continuously differentiable of all orders up to p for all* $(z,\theta) \in \mathbb{T} \times \Theta$.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Then $\phi_{\boldsymbol{L}}(\theta) = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M}(\theta))})^{\dagger}\boldsymbol{L})\boldsymbol{M}(\theta)^{\dagger}\boldsymbol{N}(\theta)I_n$ *is continuously differentiable of order* p *with respect to* θ *in the* μ*-essential supremum norm.*

Proof. See Appendix D. □

Condition (vii) of Theorem 8 is a direct strengthening of condition (iv) of Theorem 7. When $M(z,\theta)$ and $N(z,\theta)$ are Laurent matrix polynomials of uniformly bounded degree, as is usually the case in macroeconomic models, the pth-order continuous differentiability of the coefficients of $M(z,\theta)$ and $N(z,\theta)$ is sufficient to ensure this condition is satisfied.

To see Theorem 8 in action, consider the regularized solution to the example from the previous subsection. This can be obtained analytically in just a handful of steps.

$$\begin{aligned}\phi_{\boldsymbol{I}}(\theta) &= \boldsymbol{M}(\theta)^{*}(\boldsymbol{M}(\theta)\boldsymbol{M}(\theta)^{*})^{-1}I_2 \\ &= \begin{bmatrix} \boldsymbol{V}^{(-2)} & \theta\boldsymbol{V}^{(-1)} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\left(\begin{bmatrix} \boldsymbol{V}^{2} & \boldsymbol{0} \\ \theta\boldsymbol{V} & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{V}^{(-2)} & \theta\boldsymbol{V}^{(-1)} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{V}^{(-2)} & \theta\boldsymbol{V}^{(-1)} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{I} & \theta\boldsymbol{V} \\ \theta\boldsymbol{V}^{(-1)} & (\theta^2+1)\boldsymbol{I} \end{bmatrix}^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.\end{aligned}$$

To invert the operator in the expression above, we note that it is of the form defined in Definition 7, with underlying $\mathcal{W}^{2\times 2}$ element, $\begin{bmatrix} 1 & \theta z \\ \theta z^{-1} & (\theta^2+1) \end{bmatrix}$. This matrix has the Wiener–Hopf factorization $\begin{bmatrix} 1 & \frac{\theta}{\theta^2+1}z \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \frac{1}{\theta^2+1} & 0 \\ \theta z^{-1} & \theta^2+1 \end{bmatrix}$. Thus,

$$\begin{aligned}&\phi_{\boldsymbol{I}}(\theta) \\ &= \begin{bmatrix} \boldsymbol{V}^{(-2)} & \theta\boldsymbol{V}^{(-1)} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\left(\begin{bmatrix} \boldsymbol{I} & \frac{\theta}{\theta^2+1}\boldsymbol{V} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \frac{1}{\theta^2+1}\boldsymbol{I} & \boldsymbol{0} \\ \theta\boldsymbol{V}^{(-1)} & (\theta^2+1)\boldsymbol{I} \end{bmatrix}\right)^{-1}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{V}^{(-2)} & \theta\boldsymbol{V}^{(-1)} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\left(\begin{bmatrix} (\theta^2+1)\boldsymbol{I} & \boldsymbol{0} \\ -\theta\boldsymbol{V}^{(-1)} & \frac{1}{\theta^2+1}\boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{I} & -\frac{\theta}{\theta^2+1}\boldsymbol{V} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}\right)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{V}^{(-2)} & -\frac{\theta}{1+\theta^2}\boldsymbol{V}^{(-2)}\boldsymbol{V} + \frac{\theta}{1+\theta^2}\boldsymbol{V}^{(-1)} \\ -\theta\boldsymbol{V}^{(-1)} & \frac{1}{1+\theta^2}\boldsymbol{I} + \frac{\theta^2}{1+\theta^2}\boldsymbol{V}^{(-1)}\boldsymbol{V} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} z^{-2} & \frac{\theta}{1+\theta^2}z^{-1} \\ -\theta z^{-1} & \frac{1}{1+\theta^2} \end{bmatrix},\end{aligned}$$

where the last equality follows from the fact that $\ker(\boldsymbol{V}) = \mathbb{C}^{1\times 2}$ when $F = \mu I_2$. As guaranteed by Theorems 7 and 8, this solution is not just continuous as a function

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

of θ but also smooth. Note that the implicit right inverse of $\boldsymbol{M}_0$ in $\boldsymbol{M}^\dagger$ is

$$\boldsymbol{M}_-(\theta)\boldsymbol{M}(\theta)^\dagger\boldsymbol{M}_+(\theta) = \begin{cases} \begin{bmatrix} \boldsymbol{V}^{(-2)} & 0 \\ 0 & 1 \end{bmatrix}, & \theta = 0, \\ \begin{bmatrix} \boldsymbol{V}^{(-1)} & \frac{1}{1+\theta^2}\left(\boldsymbol{V}^{(-1)}\boldsymbol{V} - \boldsymbol{I}\right) \\ \boldsymbol{0} & \boldsymbol{V}^{(-1)} \end{bmatrix}, & \theta \neq 0. \end{cases}$$

Thus, it is only by using a special right inverse of $\boldsymbol{M}_0(\theta)$ that we can absorb the effect of discontinuity in the partial indices at $\theta = 0$ on the solution.

To summarize, the LREM problem in macroeconometrics is ill-posed. However, regularization produces solutions that are unique, continuous, and even smooth under very general regularity conditions.

7. APPLICATION

As an application of the spectral approach to LREMs, we apply mainstream methodology to estimate and draw inference on the nongeneric model of Section 6. We will see that the Gaussian likelihood function displays very irregular behavior, invalidating underlying assumptions of mainstream frequentist and Bayesian analysis. In turn, regularized solutions can avoid some of these anomalies.

Consider the nongeneric system of Section 6 in the time domain,

$$\begin{aligned} \boldsymbol{P}(X_{1t+2}|\mathscr{H}_t) &= \xi_{1t}, \\ \theta_0\boldsymbol{P}(X_{1t+1}|\mathscr{H}_t) + X_{2t} &= \xi_{2t}, \end{aligned} \qquad t \in \mathbb{Z},$$

with ξ an i.i.d. Gaussian sequence of mean zero and covariance matrix I_2. The objective is to estimate and draw inference on $\theta_0 \in \mathbb{R}$ from an observed dataset $x_T = (X_1^*, \ldots, X_T^*)^*$. We will follow mainstream methodology in disregarding uniqueness, identifiability, and invertibility in parametrizing the model (see Al-Sadoon and Zwiernik, 2019 for a careful parametrization that addresses all of these considerations). The solution we computed in the previous section is

$$X_t = \begin{cases} \begin{bmatrix} \xi_{1t-2} \\ \xi_{2t} \end{bmatrix}, & \theta_0 = 0, \\ \begin{bmatrix} \xi_{1t-2} + \theta_0^{-1}\xi_{2t-1} \\ -\theta_0\xi_{1t-1} \end{bmatrix}, & \theta_0 \neq 0, \end{cases} \qquad t \in \mathbb{Z}.$$

Let $\Sigma_T(\theta)$ be the covariance of x_T according to the model parametrized by θ. Then the likelihood function is

$$p(x_T|\theta) = (2\pi)^{-T}\det(\Sigma_T(\theta))^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}x_T^*\Sigma_T(\theta)^{-1}x_T\right\}.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

It is more convenient to work with

$$\ell_T(\theta) = \frac{1}{T}\log\det\Sigma_T(\theta) + \frac{1}{T}x_T^*\Sigma_T(\theta)^{-1}x_T.$$

We will refer to $\ell_T(\theta)$ as the log-likelihood function. Observe that

$$\Sigma_T(\theta) = \begin{cases} I_{2T}, & \theta = 0 \\ \begin{bmatrix} G_0 & G_1 & 0 & 0 & \cdots & 0 \\ G_1^* & G_0 & G_1 & 0 & \cdots & 0 \\ 0 & G_1^* & G_0 & G_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & G_1^* & G_0 & G_1 \\ 0 & \cdots & 0 & 0 & G_1^* & G_0 \end{bmatrix}, & \theta \neq 0, \end{cases}$$

where

$$G_0 = \begin{bmatrix} 1+\theta^{-2} & 0 \\ 0 & \theta^2 \end{bmatrix}, \qquad G_1 = \begin{bmatrix} 0 & 0 \\ -\theta & 0 \end{bmatrix}.$$

Then $\Sigma_T(0)^{-1} = I_{2T}$ and it is easily checked that for $\theta \neq 0$,

$$\Sigma_1(\theta)^{-1} = \begin{bmatrix} \frac{\theta^2}{1+\theta^2} & 0 \\ 0 & \theta^{-2} \end{bmatrix}$$

$$\Sigma_T(\theta)^{-1} = \begin{bmatrix} A & B & 0 & 0 & \cdots & 0 \\ B^* & C & B & 0 & \cdots & 0 \\ 0 & B^* & C & B & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & B^* & C & B \\ 0 & \cdots & 0 & 0 & B^* & D \end{bmatrix}, \qquad T > 1,$$

where

$$A = \begin{bmatrix} \frac{\theta^2}{1+\theta^2} & 0 \\ 0 & \frac{\theta^2+1}{\theta^2} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ \theta & 0 \end{bmatrix}, \quad C = \begin{bmatrix} \theta^2 & 0 \\ 0 & \frac{\theta^2+1}{\theta^2} \end{bmatrix},$$

$$D = \begin{bmatrix} \theta^2 & 0 \\ 0 & \theta^{-2} \end{bmatrix}.$$

It is also easily checked that $\det(\Sigma_T(\theta)) = 1+\theta^2$ for $\theta \neq 0$ for all $T \geq 1$. Substituting into the log-likelihood function, we find that

$$\ell_T(0) = \frac{1}{T}\sum_{t=1}^{T} X_t^* X_t$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

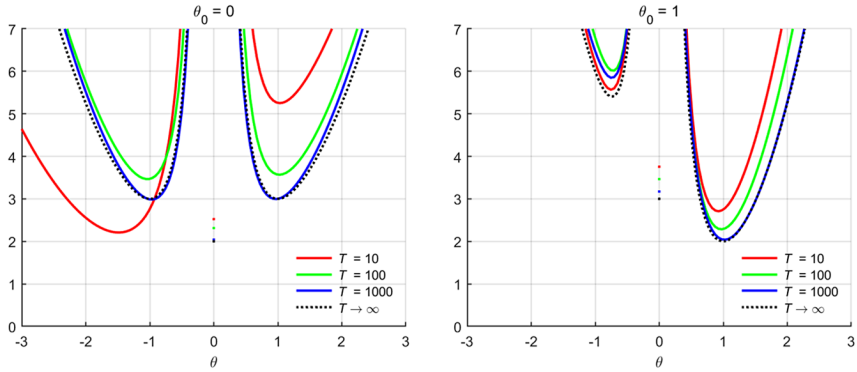


FIGURE 4. Log-likelihood of non-regularized solutions to the nongeneric model.

and, for $\theta \neq 0$,

$$\ell_T(\theta) = \begin{cases} \log(1+\theta^2) + \frac{\theta^2}{1+\theta^2} X_{11}^2 + \theta^{-2} X_{21}^2, & T = 1, \\ \frac{1}{T}\log(1+\theta^2) + \frac{1}{T} X_1^* A X_1 + \frac{1}{T}\sum_{t=2}^{T-1} X_t^* C X_t \\ \quad + \frac{2}{T}\sum_{t=1}^{T-1} X_t^* B X_{t+1} + \frac{1}{T} X_T^* D X_T, & T > 1. \end{cases}$$

It is clear that

$$\lim_{0 \neq |\theta| \to 0} \ell_T(\theta) = \lim_{|\theta| \to \infty} \ell_T(\theta) = \infty, \qquad T \geq 1,$$

almost surely because x_T has a continuous distribution regardless of the value of θ_0. Since $\ell_T(\theta)$ is continuous on $\mathbb{R}\backslash\{0\}$ and bounded from below, $\ell_T(\theta)$ must attain local minima somewhere to the left and to the right of $\theta = 0$ with probability 1. Thus, almost surely, $\ell_T(\theta)$ is W shaped, diverging as $0 \neq |\theta| \to 0$ and as $|\theta| \to \infty$, although it has a finite value at $\theta = 0$. See Figure 4 for an illustration with simulated data. It follows that the likelihood function $p(x_T|\theta)$ almost surely has at least three local maxima, one of which is $\theta = 0$. Moreover, almost surely, $\lim_{0 \neq |\theta| \to 0} p(x_T|\theta) = 0$ and $p(x_T|\theta)|_{\theta=0} > 0$, so the likelihood function almost surely has a simple discontinuity at $\theta = 0$.

Taking the limit $T \to \infty$ is straightforward as most terms vanish. By the strong law of large numbers, almost surely,

$$\lim_{T \to \infty} \ell_T(\theta) = \begin{cases} \operatorname{tr}(\mathbb{E}(X_0 X_0^*)), & \theta = 0, \\ \operatorname{tr}(C\mathbb{E}(X_0 X_0^*) + 2B\mathbb{E}(X_1 X_0^*)), & \theta \neq 0. \end{cases}$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

The moments above can be read from G_0 and G_1 evaluated at $\theta = \theta_0$, so the right-hand side is

$$\ell(\theta,\theta_0) = \begin{cases} 2, & \theta = 0, \theta_0 = 0, \\ \theta_0^2 + 1 + \theta_0^{-2}, & \theta = 0, \theta_0 \neq 0, \\ \theta^2 + 1 + \theta^{-2}, & \theta \neq 0, \theta_0 = 0, \\ \theta^2(1+\theta_0^{-2}) + (1+\theta^{-2})\theta_0^2 - 2\theta\theta_0, & \theta \neq 0, \theta_0 \neq 0. \end{cases}$$

See Figure 4 for plots of $\ell(\theta,0)$ and $\ell(\theta,1)$. It is easy to check that $\ell(\theta,0)$ has local minima at 0, $+1$, and -1 and if $\theta_0 \neq 0$, $\ell(\theta,\theta_0)$ has local minima at 0, θ_0, and a third point of the opposite sign to θ_0 (the expression is complicated so we omit it). Note that for $\theta \in \mathbb{R}\backslash\{0\}$, $\ell_T(\theta) - \ell(\theta,\theta_0)$ is expressible as a sum of eight terms, each of which is a product of a continuous function in θ and a quantity that converges almost surely to zero. Thus, $\lim_{T\to\infty} \sup_{\theta\in\Theta} |\ell_T(\theta) - \ell(\theta,\theta_0)| = 0$ for any compact subset $\Theta \subset \mathbb{R}\backslash\{0\}$.

Consider the case $\theta_0 = 0$. It is clear from the discussion above that the maximum likelihood estimator is consistent, although it is not asymptotically normal, and any prior that has an atom at the point $\theta = 0$ will yield a posterior distribution that concentrates at the true value. However, this analytical approach cannot be carried over to a general methodology because analytical descriptions of solutions are generally infeasible for all but the simplest LREMs; the mapping from (M,N) to any solution is highly nonlinear. The practitioner who seeks to estimate this model on a computer may suspect a discontinuity at $\theta = 0$ because they will be able to see the solution behaving erratically near that point but being short of an analytical description of the solution: (i) they will not have any theoretical guarantees as to what happens when $\theta \to 0$ because none exist and (ii) they will not be able to compute the solution at $\theta = 0$ reliably due to the numerical instability of computing Wiener–Hopf factorizations of nongeneric elements of $\mathcal{W}^{m\times m}$.

Now consider the likely outcome of applying mainstream econometric methodology to the data when $\theta_0 = 0$.

Let us begin with the most basic of empirical analyses: plotting the likelihood function. Each evaluation of the likelihood function is obtained by solving the model numerically then using the Kalman filter (Canova, 2011; DeJong and Dave, 2011; Herbst and Schorfheide, 2016). Without knowing a priori the importance of the likelihood function at $\theta = 0$, the researcher will not know to include that point specifically in the list plot of the likelihood function. Even if they did include it, the plot is not guaranteed to reveal the value of the likelihood function at this point because of the numerical instability of the computation at this point. Thus, the plot of the likelihood function will appear to have modes near ± 1, while the population parameter, $\theta_0 = 0$, may (and most likely will) appear to be the least likely point of the parameter space because $\lim_{0\neq|\theta|\to 0} p(x_T|\theta) = 0$.

Mainstream frequentist analysis utilizes variations of the Newton–Raphson algorithm to minimize $\ell_T(\theta)$, then obtains confidence intervals from derivatives of $\ell_T(\theta)$ (Canova, 2011; DeJong and Dave, 2011). This exercise is almost certain

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

to fail in our setting because, again, without knowing a priori that the point $\theta = 0$ is crucial to the analysis, the algorithm will be initialized in $\mathbb{R}\backslash\{0\}$ and it will converge almost surely to (approximately) either $+1$ or -1 depending on where it is initialized. The standard error of the estimate, computed using the second derivative of $\ell_T(\theta)$ at the estimated θ, will be of order $O(T^{-1/2})$ almost surely so as the sample size gets larger it will appear to the researcher that the estimate is precise when it is in fact completely wrong.

Mainstream Bayesian analysis utilizes variations on the random walk Metropolis–Hastings algorithm to sample from the posterior distribution function,

$$p(\theta|x_T) \propto p(x_T|\theta)p(\theta),$$

where $p(\theta)$ is a prior distribution function, before reporting estimates of the mean or median of the posterior distribution as well as credibility regions (Canova, 2011; DeJong and Dave, 2011; Herbst and Schorfheide, 2016). Here, again, estimation is certain to fail because, being unaware of the significance of the point $\theta = 0$, the researcher is likely to use a continuous prior that places probability zero at the set $\{0\}$. The result is that the posterior distribution will concentrate around ± 1. The only way to allow the prior to be informative about θ is to allow for an atom in the prior at $\theta = 0$, but none of the aforementioned textbooks present algorithms that can accommodate posterior distributions with atoms.

In contrast to the above, the regularized solution,

$$\tilde{X}_{1t} = \xi_{1t-2} + \frac{\theta_0}{1+\theta_0^2}\xi_{2t-1},$$
$$\tilde{X}_{2t} = -\theta_0\xi_{1t-1} + \frac{1}{1+\theta_0^2}\xi_{2t},$$

displays no such anomalies. If $\tilde{x}_T = (\tilde{X}_1^*, \ldots, \tilde{X}_T^*)^*$ and $\tilde{\Sigma}_T(\theta)$ is the covariance of $\tilde{x}_T$ according to the model parametrized by θ, then the ijth block of $\tilde{\Sigma}_T(\theta)$ is of the form $\int z^{i-j}\boldsymbol{\phi_I}(\theta)\boldsymbol{\phi_I}(\theta)^* d\mu$. Since $\boldsymbol{\phi_I}(\theta)$ is continuously differentiable of any order with respect to θ in the μ-essential supremum norm, we can exchange the integration with differentiation with respect to θ, so $\tilde{\Sigma}_T(\theta)$ is continuously differentiable of any order with respect to θ. This implies that the same is true of the likelihood function at every point $\theta \in \mathbb{R}$ where $\tilde{\Sigma}_T(\theta)^{-1}$ exists. It is easy to check that $\tilde{\Sigma}_T(\theta)$ and $\tilde{\Sigma}_T(\theta)^{-1}$ have the same block structure as $\Sigma_T(\theta)$ and $\Sigma_T(\theta)^{-1}$, respectively. The only difference is that now

$$G_0 = \begin{bmatrix} \frac{\theta^2}{(\theta^2+1)^2} + 1 & 0 \\ 0 & \theta^2 + \frac{1}{(\theta^2+1)^2} \end{bmatrix}, \quad G_1 = \begin{bmatrix} 0 & 0 \\ \theta\left(\frac{1}{(\theta^2+1)^2} - 1\right) & 0 \end{bmatrix},$$
$$A = \begin{bmatrix} \frac{(\theta^2+1)^2}{\theta^4+3\theta^2+1} & 0 \\ 0 & \frac{\theta^4+3\theta^2+1}{(\theta^2+1)^2} \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ \frac{\theta^3(\theta^2+2)}{(\theta^2+1)^2} & 0 \end{bmatrix},$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

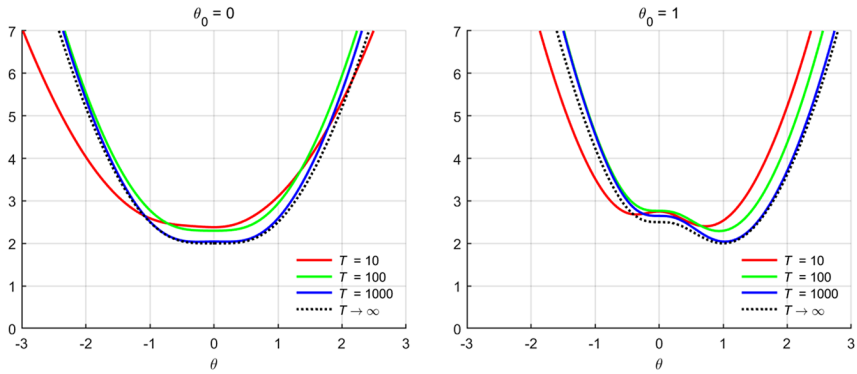


FIGURE 5. Log-likelihood of regularized solutions to the nongeneric model.

$$C = \begin{bmatrix} \frac{\theta^6+2\theta^4+\theta^2+1}{(\theta^2+1)^2} & 0 \\ 0 & \frac{\theta^4+3\theta^2+1}{(\theta^2+1)^2} \end{bmatrix}, \qquad D = \begin{bmatrix} \frac{\theta^6+2\theta^4+\theta^2+1}{(\theta^2+1)^2} & 0 \\ 0 & \frac{(\theta^2+1)^2}{\theta^4+3\theta^2+1} \end{bmatrix}.$$

Figure 5 provides a plot of the log-likelihood function of the regularized solution with simulated data along with the almost sure limit in T. It is clear now that frequentist and Bayesian analyses can be carried out using the regularized solution. Note that the first, second, and third derivatives of the limiting log-likelihood function vanish at $\theta = 0$ when $\theta_0 = 0$. Thus, a slightly more delicate frequentist analysis is necessary (see, e.g., Rotnitzky et al., 2000). This is, again, made possible by the existence of derivatives of all orders, thanks to regularization.

8. CONCLUSION

This paper has extended the LREM literature in the direction of spectral analysis. It has done so by relaxing common assumptions and developing a new regularized solution. The spectral approach has allowed us to study examples of limiting Gaussian likelihood functions of simple LREMs, which demonstrate the advantages of the new regularized solution as well as highlighting weaknesses in mainstream methodology. For the remainder, we consider some implications for future work.

The regularized solution proposed in this paper is the natural one to consider for the frequency domain. However, its motivation has been entirely econometric in nature and this begs the question of whether it can be derived from decision-theoretic foundations as proposed by Taylor (1977). Regularization has already made inroads into decision theory (Gabaix, 2014). This line of inquiry may yield other forms of regularization which may have more interesting dynamic or statistical properties.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

The parameter space in LREMs can be disconnected, and it can matter a great deal where one initializes their optimization routine to find the maximum likelihood estimator or their exploration routine for sampling from the posterior distribution. Therefore, it would be useful to develop simple preliminary estimators of LREMs analogous to the results for VARMA (e.g., Sections 8.4 and 11.5 of Brockwell and Davis, 1991) that can provide good initial conditions for frequentist and Bayesian algorithms.

Wiener–Hopf factorization theory has been demonstrated here and in previous work (Onatski, 2006; Al-Sadoon, 2018; Al-Sadoon and Zwiernik, 2019) to be the appropriate mathematical framework for analyzing LREMs. This begs the question of what is the appropriate framework for nonlinear rational expectations models. The hope is that the mathematical insights from the theory of linear models will allow for important advances in nonlinear modeling and inference.

Finally, researchers often rely on high-level assumptions as tentative placeholders when a result seems plausible but a proof from first principles is not apparent. Continuity of solutions to LREMs with respect to parameters has for a long time been one such high-level assumption in the LREM literature. The fact that it is generally false should give us pause to reflect on the prevalence of this technique. At the same time, the author hopes to have conveyed a sense of optimism that theoretical progress from first principles is possible.

APPENDICES

A. Parametrizing the Set of Solutions

By Lemma 2, the set of solutions to (13) when $\kappa_m \geq 0$ is the affine space $\boldsymbol{M}^{(-1)}\boldsymbol{N}I_n + \ker(\boldsymbol{M})$, where

$$\ker(\boldsymbol{M}) = \boldsymbol{M}_-^{-1}\ker(\boldsymbol{M}_0).$$

It suffices to parametrize

$$\ker(\boldsymbol{M}_0) = \left\{ \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_m \end{bmatrix} : \psi_i \in \ker(\boldsymbol{V}^{\kappa_i}),\ i = 1, \ldots, m \right\}.$$

The problem then reduces to the parametrization of $\ker(\boldsymbol{V}^{\kappa})$ for $\kappa \geq 0$ and, by Lemma 1(viii), the problem reduces even further to the parametrization of $\ker(\boldsymbol{V})$. No, if $\ker(\boldsymbol{V}) = \{0\}$, then $\ker(\boldsymbol{V}^{\kappa}) = \{0\}$, for all $\kappa \geq 0$, and so $\ker(\boldsymbol{M}_0) = \{0\}$. If $\ker(\boldsymbol{V}) \neq \{0\}$, we may find an orthonormal basis for it,

$$\Upsilon_1, \ldots, \Upsilon_r \in \ker(\boldsymbol{V}).$$

Note that $r \leq n$ by Lemma 1(vi). Lemma 1(viii) then implies that

$$\Upsilon_1, \ldots, \Upsilon_r, \boldsymbol{V}^{(-1)}\Upsilon_1, \ldots, \boldsymbol{V}^{(-1)}\Upsilon_r, \ldots, \boldsymbol{V}^{(1-\kappa)}\Upsilon_1, \ldots, \boldsymbol{V}^{(1-\kappa)}\Upsilon_r$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

is an orthonormal basis for $\ker(\boldsymbol{V}^{\kappa})$ for $\kappa \geq 0$. Thus,

$$\ker(\boldsymbol{M}_0) = \left\{ K\Upsilon : K_{ij}(z) = \sum_{s=0}^{\kappa_i - 1} K_{sij} z^{-s}, K_{sij} \in \mathbb{C},\ i = 1, \ldots, m,\ j = 1, \ldots, r,\ s = 0, \ldots, \kappa_i - 1 \right\},$$

where

$$\Upsilon = \begin{bmatrix} \Upsilon_1 \\ \vdots \\ \Upsilon_r \end{bmatrix}.$$

Therefore, $\ker(\boldsymbol{M}_0)$ is the space obtained by pre-multiplying Υ by the space of $m \times n$ complex matrix polynomials in z^{-1} with ith row degrees bounded by $\kappa_i - 1$. Notice that the rows of K associated with partial indices at zero are identically zero. The dimension of $\ker(\boldsymbol{M})$ can be obtained by counting the coefficients K_{sij} above, $\dim\ker(\boldsymbol{M}) = r(\kappa_1 + \cdots + \kappa_m)$.

B. Relation to Previous Literature

This section provides a review of the approaches of Whiteman (1983), Onatski (2006), Tan and Walker (2015), Tan (2019), and Meyer-Gohde and Tzaawa-Krenzler (2023) (henceforth, the previous literature). We will see that the previous literature makes significantly stronger assumptions about ξ that also make it cumbersome to address the topic of zeros of $\det(M)$. On the other hand, the stronger assumptions do allow for closed-form expressions of solutions.

The previous literature takes as its starting point the existence of a Wold decomposition for ξ with no purely deterministic part. Thus, it imposes that F be absolutely continuous with respect to μ and that the spectral density matrix has an analytic spectral factorization with spectral factors of fixed rank μ–a.e. (Rozanov, 1967, Thm. II.4.1). The conditions are

$$\begin{gathered} dF = \Gamma\Gamma^* d\mu, \\ \sum_{s=0}^{\infty} \Gamma_s z^{-s} \text{ converges for } |z| > 1, \text{ where } \Gamma_s = \int z^s \Gamma d\mu, \\ \operatorname{rank}(\Gamma(z)) = r, \quad \mu - \text{a.e. } z \in \mathbb{T}. \end{gathered} \tag{A.1}$$

Theorem II.8.1 of Rozanov (1967) provides equivalent analytical conditions. This paper has demonstrated that a complete spectral theory of LREMs is possible without imposing any such restrictions.

Clearly, the spectral factorization in (A.1) is not unique. However, there always exists a spectral factor Γ unique up to right multiplication by a unitary matrix such that there is an $\Upsilon \in H_0^r$ satisfying

$$\Upsilon(z)\Gamma(z) = I_r, \qquad \mu - \text{a.e. } z \in \mathbb{T}$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

(Rozanov, 1967, Sect. II.4). This choice of spectral factor yields a Wold representation. To see this, let

$$\zeta_t = \int z^t \Upsilon d\Phi, \qquad t \in \mathbb{Z}.$$

Then

$$\boldsymbol{E}\zeta_t \zeta_s^* = \int z^{t-s} \Upsilon dF \Upsilon^* = \int z^{t-s} \Upsilon \Gamma \Gamma^* \Upsilon^* d\mu = \int z^{t-s} I_r d\mu = \begin{cases} I_r, & t = s, \\ 0, & t \neq s. \end{cases}$$

Thus, ζ is an r-dimensional standardized white noise process causal in ξ. It follows easily that $\ker(\boldsymbol{V}) = \{x^* \Upsilon : x \in \mathbb{C}^r\}$ so that $\dim \ker(\boldsymbol{V}) = r$. Since Υ is a left inverse of Γ,

$$\|\Gamma\Upsilon - I_n\|_{H^n}^2 = \operatorname{tr}\left(\int (\Gamma\Upsilon - I_n)\Gamma\Gamma^*(\Gamma\Upsilon - I_n)^* d\mu \right) = 0$$

and so

$$\xi_t = \int z^t d\Phi = \int z^t \Gamma\Upsilon d\Phi = \int \Gamma z^t \Upsilon d\Phi = \sum_{s=0}^{\infty} \Gamma_s \zeta_{t-s}, \quad t \in \mathbb{Z},$$

is a Wold representation with

$$\Gamma_s = \int z^s \Gamma d\mu, \qquad s \geq 0.$$

The previous literature imposes a priori the representation

$$X_t = \sum_{s=0}^{\infty} \Xi_s \zeta_{t-s}, \qquad t \in \mathbb{Z}$$

before attempting to solve for the coefficients Ξ_s by complex analytic methods (Whiteman, 1983; Tan and Walker, 2015; Tan, 2019; Meyer-Gohde and Tzaawa-Krenzler, 2023) or by Wiener–Hopf factorization (Onatski, 2006). This method cannot yield the correct solution if ξ has a nontrivial purely deterministic part. In fact, the representation above need not be assumed a priori and can be derived as a consequence of Theorem 3 under conditions (A.1).

$$\begin{aligned} X_t &= \int z^t \left(\boldsymbol{M}^{(-1)} \boldsymbol{N} I_n + \boldsymbol{M}_-^{-1} \psi \right) d\Phi \\ &= \int z^t \left(\boldsymbol{M}^{(-1)} \boldsymbol{N} I_n + \boldsymbol{M}_-^{-1} \psi \right) \Gamma\Upsilon d\Phi \\ &= \int \Xi z^t \Upsilon d\Phi, \qquad t \in \mathbb{Z}, \end{aligned}$$

where

$$\Xi = \left(\boldsymbol{M}^{(-1)} \boldsymbol{N} I_n + \boldsymbol{M}_-^{-1} \psi \right) \Gamma.$$

Thus, we have obtained the spectral characteristic of X relative to the random measure associated with ζ. It follows that X is indeed representable as a moving average in ζ with coefficients,

$$\Xi_s = \int z^s \Xi d\mu, \qquad s \geq 0.$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

It is important to note, however, that the stronger assumptions of the previous literature lead to a more explicit expression for spectral characteristics of solutions. In particular,

$$\begin{aligned}\boldsymbol{NI}_n &= \boldsymbol{P}(N|H_0^m)\\ &= \boldsymbol{P}(N\Gamma\Upsilon|H_0^m)\\ &= [N\Gamma]_-\Upsilon.\end{aligned}$$

This follows from the fact that $\{z^t\Upsilon : t > 0\}$ is orthogonal to H_0^m. It then follows that

$$\begin{aligned}\boldsymbol{M}^{(-1)}\boldsymbol{NI}_n + \boldsymbol{M}_-^{-1}\psi &= \boldsymbol{M}_-^{-1}\boldsymbol{M}_0^{(-1)}\boldsymbol{M}_+^{-1}([N\Gamma]_-\Upsilon) + \boldsymbol{M}_-^{-1}\psi\\ &= M_-^{-1}M_0^{-1}\boldsymbol{P}(M_+^{-1}[N\Gamma]_-\Upsilon|H_0^m) + M_-^{-1}\psi\\ &= M_-^{-1}M_0^{-1}[M_+^{-1}[N\Gamma]_-]_-\Upsilon + M_-^{-1}\psi\\ &= M_-^{-1}M_0^{-1}[M_+^{-1}N\Gamma]_-\Upsilon + M_-^{-1}\psi,\end{aligned}$$

because $[M_+^{-1}[N\Gamma]_-]_- = [M_+^{-1}N\Gamma]_- - [M_+^{-1}[N\Gamma]_+]_- = [M_+^{-1}N\Gamma]_- - 0$. Finally, using the results of Appendix 1, $\psi = K\Upsilon$, where K is a matrix polynomial in z^{-1}. Thus,

$$\phi = M_-^{-1}M_0^{-1}[M_+^{-1}N\Gamma]_-\Upsilon + M_-^{-1}K\Upsilon$$

and so

$$\Xi = M_-^{-1}M_0^{-1}[M_+^{-1}N\Gamma]_- + M_-^{-1}K.$$

An interesting special case is when M, N, and Γ are rational. If M is rational, any Wiener–Hopf factors $M_\pm$ are rational as well (Clancey and Gohberg, 1981, Thm. I.2.1). If Γ is rational, the associated Υ can also be chosen to be rational (Baggio and Ferrante, 2016, Thm. 1). Thus, ϕ and Ξ are also rational.

Finally, we note that the previous literature has avoided any mention of zeros of $\det(M)$. Indeed, it is substantially more difficult to deal with zeros of $\det(M)$ without the general theory of this paper because one no longer has access to degenerate spectral measures (e.g., the Dirac measure) that can straightforwardly excite the instability of the system.

C. Solving the Nongeneric System

Consider solving the system (18) by hand when $\theta \neq 0$. In the time domain, this is given by

$$\boldsymbol{P}(X_{1t+2}|\mathscr{H}_t) = \xi_{1t}, \tag{A.2}$$

$$\theta\boldsymbol{P}(X_{1t+1}|\mathscr{H}_t) + X_{2t} = \xi_{2t}, \tag{A.3}$$

where ξ is a standard white noise process. Macroeconomic textbooks (e.g., Sargent, 1979) implicitly assume the admissibility of the following elementary operations for solving LREMs.

(i) Multiply both sides of equation i by a nonzero constant.
(ii) Apply the mapping, $h \mapsto \boldsymbol{P}(\boldsymbol{U}^s h, |\mathscr{H}_t)$, to both sides of equation $j \neq i$ and add the resultant to equation i.
(iii) Permute equations i and j.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

These elementary operations are a direct generalization of the familiar row-reduction elementary operations in linear algebra ($s = 0$) as well as the elementary operations for VARMA manipulation ($s \leq 0$). See page 39 of Hannan and Deistler (1988).

We seek a process X causal in ξ that solves (A.2) and (A.3). The most immediate solution does not require any of the elementary operations above. It can be obtained by noting that, since $\boldsymbol{P}(\xi_{1t}|\mathscr{H}_t) = \xi_{1t}$, we can set

$$X_{1t} = \xi_{1t-2}$$

then substitute into equation (A.3) and, noting that $\boldsymbol{P}(\xi_{1t-1}|\mathscr{H}_t) = \xi_{1t-1}$, we obtain

$$\theta\xi_{1t-1} + X_{2t} = \xi_{2t}.$$

We therefore obtain the solution

$$X_{1t} = \xi_{1t-2},$$
$$X_{2t} = -\theta\xi_{1t-1} + \xi_{2t},$$

which is clearly causal in ξ. However, this is not the only way of solving the problem. If we multiply both sides of equation (A.2) by $-\theta$ (operation (i)), we obtain

$$-\theta\boldsymbol{P}(X_{1t+2}|\mathscr{H}_t) = -\theta\xi_{1t}.$$

Now apply $h \mapsto \boldsymbol{P}(\boldsymbol{U}h|\mathscr{H}_t)$ to both sides of equation (A.3) and add the resultant to equation (A.2) (operation (ii)) to obtain

$$\boldsymbol{P}(X_{2t+1}|\mathscr{H}_t) = -\theta\xi_{1t}.$$

Thus, we can set

$$X_{2t} = -\theta\xi_{1t-1}.$$

Plugging this into (A.3), we obtain

$$\theta\boldsymbol{P}(X_{1t+1}|\mathscr{H}_t) - \theta\xi_{1t-1} = \xi_{2t},$$

which allows us to set (using operation (i)),

$$X_{1t} = \xi_{1t-2} + \theta^{-1}\xi_{2t-1}.$$

Finally, permuting the two equations we have obtained (operation (iii)), we obtain the alternative solution

$$X_{1t} = \xi_{1t-2} + \theta^{-1}\xi_{2t-1},$$
$$X_{2t} = -\theta\xi_{1t-1}.$$

This is precisely the solution obtained in Section 6.

D. Proofs

Theorems 7 and 8 require some additional technical results that we need to develop first.

Define for $\phi \in H^m$ and $\omega \in \mathbb{R}$, $\boldsymbol{S}_\omega\phi(z) = \phi(ze^{\mathrm{i}\omega})$ and, for $\omega \neq 0$, $\boldsymbol{\Delta}_\omega\phi = \frac{1}{\omega}(\boldsymbol{S}_\omega\phi - \phi)$. Set $\boldsymbol{d}\phi$ to be the H^m limit of $\boldsymbol{\Delta}_\omega\phi$ as $\omega \to \infty$ whenever it exists. Clearly, $\boldsymbol{S}_\omega$ is a unitary operator on H_0^m, $\boldsymbol{\Delta}_\omega$ is a bounded operator on H_0^m, and $\boldsymbol{d}$ is not generally bounded on H_0^m.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

We will need the following inequality, inspired by similar inequalities of Anderson (1985) and Green and Anderson (1987), to prove our main results.

LEMMA 6. *If $F = \mu I_n$ and $\phi, \boldsymbol{d}\phi \in H^m$, then*

$$\|\phi\|_\infty \leq \|\phi\|_{H^m} + \|\boldsymbol{d}\phi\|_{H^m}.$$

Proof. Let $\tau(v,w)$ be the counterclockwise segment of $\mathbb{T}$ from $v \in \mathbb{T}$ to $w \in \mathbb{T}$, and let $1_{\tau(v,w)}$ be the indicator function for that segment. By a change of variables,

$$\int 1_{\tau(v,w)} \boldsymbol{\Delta}_\omega \phi d\mu = \frac{1}{\omega} \int_{\tau(w, we^{i\omega})} \phi d\mu - \frac{1}{\omega} \int_{\tau(v, ve^{i\omega})} \phi d\mu.$$

Since $\phi \in H^m$, ϕ is μ-integrable, so the right-hand side converges to $\phi(w) - \phi(v)$ for μ – a.e. w and v, the Lebesgue points of ϕ (Rudin, 1986, Theorem 7.10). On the other hand, since $\boldsymbol{\Delta}_\omega \phi$ converges in H^m, the continuity of the inner product implies that the left-hand side converges to $\int 1_{\tau(v,w)} \boldsymbol{d}\phi d\mu$. Therefore,

$$\int 1_{\tau(v,w)} \boldsymbol{d}\phi d\mu = \phi(w) - \phi(v), \qquad \mu - \text{a.e. } w, v \in \mathbb{T}.$$

It follows by the triangle and Jensen's inequality that

$$\|\phi(w)\|_{\mathbb{C}^{m\times n}} \leq \|\phi(v)\|_{\mathbb{C}^{m\times n}} + \int 1_{\tau(v,w)} \|\boldsymbol{d}\phi\|_{\mathbb{C}^{m\times n}} d\mu, \qquad \mu - \text{a.e. } w, v \in \mathbb{T}.$$

Now, among all Lebesgue points of ϕ, choose a v such that $\|\phi(v)\|^2_{\mathbb{C}^{m\times n}} \leq \|\phi\|^2_{H^m}$. Thus,

$$\|\phi(w)\|_{\mathbb{C}^{m\times n}} \leq \|\phi\|_{H^m} + \int \|\boldsymbol{d}\phi\|_{\mathbb{C}^{m\times n}} d\mu, \qquad \mu - \text{a.e. } w \in \mathbb{T}.$$

Since $\int \|\boldsymbol{d}\phi\|_{\mathbb{C}^{m\times n}} d\mu \leq \left(\int \|\boldsymbol{d}\phi\|^2_{\mathbb{C}^{m\times n}} d\mu \right)^{1/2} = \|\boldsymbol{d}\phi\|_{H^m}$, we have

$$\|\phi(w)\|_{\mathbb{C}^{m\times n}} \leq \|\phi\|_{H^m} + \|\boldsymbol{d}\phi\|_{H^m}, \qquad \mu - \text{a.e. } w \in \mathbb{T}.$$

Now simply take the μ-essential supremum on the left-hand side (Rudin, 1986, p. 66). □

We will also need a notion of differentiation of operators that interacts well with $\boldsymbol{d}$. For an arbitrary $\boldsymbol{A} : H^m \to H^l$, define $\boldsymbol{dA}$ to be the operator limit of $\frac{1}{\omega}(\boldsymbol{S}_\omega \boldsymbol{A} \boldsymbol{S}_{-\omega} - \boldsymbol{A})$ as $\omega \to 0$ if it exists. It is easily shown that for $\phi \in H^m$ and $\boldsymbol{B} : H^n \to H^m$,

$$\boldsymbol{d}(\boldsymbol{A}\phi) = \boldsymbol{dA}\phi + \boldsymbol{A}\boldsymbol{d}\phi, \qquad \boldsymbol{d}(\boldsymbol{AB}) = \boldsymbol{dAB} + \boldsymbol{AdB},$$

whenever $\boldsymbol{d}\phi$, $\boldsymbol{dA}$, and $\boldsymbol{dB}$ exist.

If $M : \mathbb{C} \to \mathbb{C}^{m\times n}$ is analytic in a neighborhood of $\mathbb{T}$, $\boldsymbol{dM}$ takes a particularly simple form.

LEMMA 7. *Let $M : \mathbb{C} \to \mathbb{C}^{m\times n}$ be analytic in neighborhood of $\mathbb{T}$ and let M' be its derivative, then $\boldsymbol{dM} : \phi \mapsto \boldsymbol{P}\left(\mathrm{i}zM'\phi \middle| H^m_0\right)$.*

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Proof. Since $\mathrm{i}zM'$ is also analytic in a neighborhood of $\mathbb{T}$ (Rudin, 1986, Cor. 10.16), its restriction to $\mathbb{T}$ is in $\mathcal{W}^{m\times n}$ and $\mathrm{i}z\boldsymbol{M'}$ is well defined on H_0^n. Next, since $\boldsymbol{S}_\omega$ is a unitary operator on H_0^m,

$$\begin{aligned}\frac{1}{\omega}(\boldsymbol{S}_\omega \boldsymbol{M}\boldsymbol{S}_{-\omega}-\boldsymbol{M})\phi &= \frac{1}{\omega}\left(\boldsymbol{P}\left(\boldsymbol{S}_\omega(\boldsymbol{M}\boldsymbol{S}_{-\omega}\phi)|\boldsymbol{S}_\omega H_0^m\right)-\boldsymbol{P}\left(M\phi|H_0^m\right)\right)\\ &= \frac{1}{\omega}\left(\boldsymbol{P}\left((\boldsymbol{S}_\omega M)\phi|H_0^m\right)-\boldsymbol{P}\left(M\phi|H_0^m\right)\right)\\ &= \boldsymbol{P}\left((\boldsymbol{\Delta}_\omega M)\phi|H_0^m\right),\end{aligned}$$

where we have used the fact that $\boldsymbol{S}_\omega H_0^m = H_0^m$ and the fact that $\boldsymbol{S}_\omega(\boldsymbol{M}\boldsymbol{S}_{-\omega}\phi) = (\boldsymbol{S}_\omega M)\phi$ for all $\phi \in H^n$. Since $\boldsymbol{\Delta}_\omega M - \mathrm{i}zM'$ restricted to $\mathbb{T}$ is in $\mathcal{W}^{m\times n}$, the discussion following Definition 7 implies that $\frac{1}{\omega}(\boldsymbol{S}_\omega \boldsymbol{M}\boldsymbol{S}_{-\omega}-\boldsymbol{M}) - \mathrm{i}z\boldsymbol{M'}$ is bounded in the operator norm by

$$\begin{aligned}\left\|\boldsymbol{\Delta}_\omega M - \mathrm{i}zM'\right\|_\infty &= \sup_{z\in\mathbb{T}}\left\|\frac{1}{\omega}\int_0^\omega \mathrm{i}ze^{\mathrm{i}\lambda}M'(ze^{\mathrm{i}\lambda})-\mathrm{i}zM'(z)d\lambda\right\|_{\mathbb{C}^{m\times n}}\\ &\le \sup_{z\in\mathbb{T}}\sup_{0\le\lambda\le\omega}\left\|\mathrm{i}ze^{\mathrm{i}\lambda}M'(ze^{\mathrm{i}\lambda})-\mathrm{i}zM'(z)\right\|_{\mathbb{C}^{m\times n}}.\end{aligned}$$

This converges to zero as $\omega\to 0$ by the uniform continuity of $\mathrm{i}zM'$ on $\mathbb{T}$. □

The final lemma consists of technical results, more general versions of which are due to Locker and Prenter (1980) and Callon and Groetsch (1987). Our proofs are specialized and modified so that they follow from first principles.

LEMMA 8. *Let $M \in \mathcal{W}^{m\times m}$, let $\det(M(z)) \neq 0$ for all $z \in \mathbb{T}$, and suppose the partial indices of M are nonnegative. Let $\boldsymbol{L} : H_0^m \to H^l$ be a bounded linear operator, let $\ker(\boldsymbol{M}) \cap \ker(\boldsymbol{L}) = \{0\}$, and let $\boldsymbol{W} = \boldsymbol{M}^*\boldsymbol{M} + \boldsymbol{L}^*\boldsymbol{L}$. Then the following holds:*

(i) *For $\phi, \psi \in H_0^m$, the inner product*

$$[[\phi,\psi]] = ((\boldsymbol{M}\phi,\boldsymbol{M}\psi)) + ((\boldsymbol{L}\phi,\boldsymbol{L}\psi))$$

defines a Hilbert space H_0^m and we write $\|\phi\|^2_{\mathsf{H}_0^m} = [[\phi,\phi]]$.

(ii) *H_0^m and H_0^m have equivalent norms.*

(iii) *If $\boldsymbol{A} : H_0^m \to H_0^m$, then its H_0^m adjoint is given by*

$$\boldsymbol{A}^\times = \boldsymbol{W}^{-1}\boldsymbol{A}^*\boldsymbol{W}.$$

(iv) *The H_0^m Moore–Penrose inverse of $\boldsymbol{M}$ is*

$$\boldsymbol{M}^- = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger \boldsymbol{L})\boldsymbol{M}^\dagger.$$

Proof. (i) Clearly, $[[\cdot,\cdot]]$ is an inner product and it remains to show that H_0^m is complete. Let $\{\phi_i : i = 1,2,\dots\} \subset H_0^m$ be an H_0^m Cauchy sequence. Then both $\boldsymbol{M}\phi_i$ and $\boldsymbol{L}\phi_i$ are Cauchy in H^m and H^l, respectively. They must therefore have limits $\varphi_{\boldsymbol{M}} \in H^m$ and $\varphi_{\boldsymbol{L}} \in H^l$, respectively. Now write

$$\phi_i = \phi_{i,\,\ker(\boldsymbol{M})} + \phi_{i,\,\ker(\boldsymbol{M})^\perp}, \qquad i = 1,2,\dots,$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

where $\phi_{i,\,\ker(\boldsymbol{M})} \in \ker(\boldsymbol{M})$ and $\phi_{i,\,\ker(\boldsymbol{M})^\perp} \in \ker(\boldsymbol{M})^\perp$, the orthogonal complement to $\ker(\boldsymbol{M})$ in H_0^m. Then

$$\boldsymbol{M}\phi_i = \boldsymbol{M}\phi_{i,\,\ker(\boldsymbol{M})^\perp} \to \varphi_{\boldsymbol{M}} \text{ in } H_0^m.$$

We have already established in Lemma 4 that $\boldsymbol{M}^\dagger$ is a bounded linear operator. Since $\boldsymbol{M}^\dagger\boldsymbol{M}$ is the orthogonal projection onto $\ker(\boldsymbol{M})^\perp$ in H_0^m (Groetsch, 1977, p. 47),

$$\phi_{i,\,\ker(\boldsymbol{M})^\perp} = \boldsymbol{M}^\dagger\boldsymbol{M}\phi_{i,\,\ker(\boldsymbol{M})^\perp} \to \boldsymbol{M}^\dagger\varphi_{\boldsymbol{M}} \text{ in } H_0^m.$$

It follows, since $\boldsymbol{L}$ is a bounded linear operator, that

$$\boldsymbol{L}\phi_{i,\,\ker(\boldsymbol{M})^\perp} \to \boldsymbol{L}\boldsymbol{M}^\dagger\varphi_{\boldsymbol{M}} \text{ in } H^l.$$

Therefore,

$$\boldsymbol{L}|_{\ker(\boldsymbol{M})}\phi_{i,\,\ker(\boldsymbol{M})} = \boldsymbol{L}\phi_{i,\,\ker(\boldsymbol{M})} \to \varphi_{\boldsymbol{L}} - \boldsymbol{L}\boldsymbol{M}^\dagger\varphi_{\boldsymbol{M}} \text{ in } H^l.$$

It has already been established in Lemma 2 that $\dim(\ker(\boldsymbol{M})) < \infty$; thus, $\boldsymbol{L}|_{\ker(\boldsymbol{M})}$ is of finite rank and its image is closed. It follows that $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger$ exists and is a bounded linear operator (Groetsch, 1977, Cor. 2.1.3). Moreover, $\ker(\boldsymbol{M}) \cap \ker(\boldsymbol{L}) = \{0\}$ implies that $\boldsymbol{L}|_{\ker(\boldsymbol{M})}$ is injective; therefore, the image of $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^*$ is $\ker(\boldsymbol{M})$. By Theorem 2.1.2 of Groetsch (1977), the image of $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^*$ is the image of $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger$. Thus, $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger(\boldsymbol{L}|_{\ker(\boldsymbol{M})})$ is the orthogonal projection onto $\ker(\boldsymbol{M})$ in H_0^m and so

$$\phi_{i,\,\ker(\boldsymbol{M})} = (\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger(\boldsymbol{L}|_{\ker(\boldsymbol{M})})\phi_{i,\,\ker(\boldsymbol{M})} \to (\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left(\varphi_{\boldsymbol{L}} - \boldsymbol{L}\boldsymbol{M}^\dagger\varphi_{\boldsymbol{M}}\right) \text{ in } H_0^m.$$

Therefore, ϕ_i converges in H_0^m to a point, call it ϕ_0. Finally,

$$\|\phi_i - \phi_0\|_{\mathsf{H}_0^m}^2 = \|\boldsymbol{M}(\phi_i - \phi_0)\|_{H^m}^2 + \|\boldsymbol{L}(\phi_i - \phi_0)\|_{H^l}^2$$

and the boundedness of $\boldsymbol{M}$ and $\boldsymbol{L}$ imply that ϕ_i converges to ϕ_0 in H_0^m as well.

(ii) Since $\boldsymbol{M}$ and $\boldsymbol{L}$ are bounded linear operators, there exists an upper bound $c > 0$ on their operator norms and

$$\|\phi\|_{\mathsf{H}_0^m} = \left(\|\boldsymbol{M}\phi\|_{H^m}^2 + \|\boldsymbol{L}\phi\|_{H^l}^2\right)^{1/2} \le \sqrt{2}c\|\phi\|_{H^m}, \qquad \phi \in \mathsf{H}_0^m.$$

The equivalence of $\|\cdot\|_{H^m}$ and $\|\cdot\|_{\mathsf{H}_0^m}$ on H_0^m then follows from Corollary XII.4.2 of Gohberg et al. (2003a).

(iii) For $\phi, \psi \in H_0^m$,

$$[[\phi, \psi]] = ((\boldsymbol{W}\phi, \psi)).$$

This implies that

$$((\boldsymbol{W}\boldsymbol{A}^\times\phi, \psi)) = [[\boldsymbol{A}^\times\phi, \psi]] = [[\phi, \boldsymbol{A}\psi]] = ((\boldsymbol{W}\phi, \boldsymbol{A}\psi)).$$

Since the last term is equal to $((\boldsymbol{A}^*\boldsymbol{W}\phi, \psi))$ and ϕ and ψ are arbitrary,

$$\boldsymbol{W}\boldsymbol{A}^\times = \boldsymbol{A}^*\boldsymbol{W}.$$

If $\boldsymbol{W}\phi = 0$, then $0 = ((\boldsymbol{W}\phi, \phi)) = \|\phi\|_{\mathsf{H}_0^m}^2$. This implies that $\boldsymbol{W}$ is injective. Next, if ϕ is orthogonal to the image of $\boldsymbol{W}$, then $((\phi, \boldsymbol{W}\varphi)) = 0$ for all $\varphi \in H_0^m$. Since $\boldsymbol{W}$ is self-adjoint as

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

an operator on H_0^m, we have that $((\boldsymbol{W}\phi,\varphi))=0$ for all $\varphi\in H_0^m$, in particular $0=((\boldsymbol{W}\phi,\phi))=\|\phi\|^2_{\mathsf{H}_0^m}$. Thus, $\boldsymbol{W}$ is surjective. It follows that $\boldsymbol{W}$ is invertible (Gohberg et al., 2003a, p. 283), and the expression for $\boldsymbol{A}^\times$ follows.

(iv) By the arguments used in Lemma 4, we have the following representation:

$$\begin{aligned}\boldsymbol{M}^- &= \boldsymbol{M}^\times(\boldsymbol{M}\boldsymbol{M}^\times)^{-1}\\ &= \boldsymbol{W}^{-1}\boldsymbol{M}^*\left(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}^*\right)^{-1}\\ &= \boldsymbol{W}^{-1}\boldsymbol{M}^*\left(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}^*\right)^{-1}\boldsymbol{M}\boldsymbol{M}^\dagger\\ &= \boldsymbol{W}^{-1/2}\left(\boldsymbol{W}^{-1/2}\boldsymbol{M}^*\left(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}^*\right)^{-1}\boldsymbol{M}\boldsymbol{W}^{-1/2}\right)\boldsymbol{W}^{1/2}\boldsymbol{M}^\dagger.\end{aligned}$$

The second equality follows from (iii), the third follows from Lemma 4, and the fourth follows from Theorem V.6.1 of Gohberg et al. (1990). Next, notice that the operator

$$\boldsymbol{W}^{-1/2}\boldsymbol{M}^*\left(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}^*\right)^{-1}\boldsymbol{M}\boldsymbol{W}^{-1/2}$$

is a self-adjoint projection acting on H_0^m. By Theorem II.13.1 of Gohberg et al. (2003a), it is the H_0^m orthogonal projection onto its image and it is easily seen that this is the image of $\boldsymbol{W}^{-1/2}\boldsymbol{M}^*$. Let

$$\boldsymbol{\Pi}=\boldsymbol{I}-\boldsymbol{W}^{-1/2}\boldsymbol{M}^*\left(\boldsymbol{M}\boldsymbol{W}^{-1}\boldsymbol{M}^*\right)^{-1}\boldsymbol{M}\boldsymbol{W}^{-1/2}.$$

Then $\boldsymbol{\Pi}$ is the H_0^m orthogonal projection onto $\boldsymbol{W}^{1/2}\ker(\boldsymbol{M})$, the orthogonal complement to the image of $\boldsymbol{W}^{-1/2}\boldsymbol{M}^*$ in H_0^m. Since $\ker(\boldsymbol{M})$ is the image of $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger$, $\boldsymbol{\Pi}$ is the H_0^m orthogonal projection onto the image of $\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger$. We then have that

$$\begin{aligned}\boldsymbol{\Pi} &= \left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)^\dagger\\ &= \left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\left\{\left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)^*\left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\right\}^\dagger\\ &\quad\times\left(\boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)^*\\ &= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left\{(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right\}^\dagger(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}^{1/2}\\ &= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left\{(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{L}^*\boldsymbol{L}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right\}^\dagger(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}^{1/2}\\ &= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left\{(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right\}^\dagger(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}^{1/2}\\ &= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left\{\left((\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)^*\left((\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\right\}^\dagger\\ &\quad\times(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}^{1/2}\\ &= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\left\{\left((\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\left((\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^\dagger\right)\right\}^\dagger\\ &\quad\times(\boldsymbol{L}|^*_{\ker(\boldsymbol{M})})^\dagger\boldsymbol{W}^{1/2}\end{aligned}$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

$$
\begin{aligned}
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\left\{(\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\right\}^{\dagger}(\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{W}^{1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}(\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{W}^{1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}(\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{W}^{1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}(\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{W}\boldsymbol{W}^{-1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}(\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L}|^{*}_{\ker(\boldsymbol{M})}\boldsymbol{L}\boldsymbol{W}^{-1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\left(\boldsymbol{L}|_{\ker(\boldsymbol{M})}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\right)^{*}\boldsymbol{L}\boldsymbol{W}^{-1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\left(\boldsymbol{L}|_{\ker(\boldsymbol{M})}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\right)\boldsymbol{L}\boldsymbol{W}^{-1/2}\\
&= \boldsymbol{W}^{1/2}(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L}\boldsymbol{W}^{-1/2},
\end{aligned}
$$

where we have used basic properties of the Moore–Penrose inverse (Groetsch, 1977, Sects. 2.1 and 2.2) as well as the fact that $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}$ and $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{*}$ map into $\ker(\boldsymbol{M})$. It follows that

$$
\begin{aligned}
\boldsymbol{M}^{-} &= \boldsymbol{W}^{-1/2}(\boldsymbol{I}-\boldsymbol{\Pi})\boldsymbol{W}^{1/2}\boldsymbol{M}^{\dagger}\\
&= (\boldsymbol{I}-(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L})\boldsymbol{M}^{\dagger}.
\end{aligned}
$$

□

Lemma 8(i) introduces a Hilbert space H_0^m that plays an important role in our regularization theory; H_0^m and H_0^m have the same elements but different inner products. Lemma 8(ii) implies that convergence of points in (operators on) H_0^m is equivalent to convergence of points in (operators on) H_0^m. In particular, $\boldsymbol{d}$, whether acting on points or operators, takes the same value in both spaces. On the other hand, adjoints and orthogonal projections are not the same in both spaces due to the different inner products. This is what is proven in Lemma 8(iii) and (iv).

Proof of Lemma 5. The proof that $\phi_{\boldsymbol{L}}$ solves the regularization problem is due to Callon and Groetsch (1987). We provide an alternative direct derivation. By Lemmas 2 and 4,

$$
\begin{aligned}
\min\left\{\|\boldsymbol{L}\phi\|^2_{H^l} : \boldsymbol{M}\phi = \boldsymbol{N}I_n\right\} &= \min\left\{\|\boldsymbol{L}\phi\|^2_{H^l} : \phi \in \boldsymbol{M}^{\dagger}\boldsymbol{N}I_n + \ker(\boldsymbol{M})\right\}\\
&= \min\left\{\|\boldsymbol{L}(\boldsymbol{M}^{\dagger}\boldsymbol{N}I_n + \chi)\|^2_{H^l} : \chi \in \ker(\boldsymbol{M})\right\}\\
&= \min\left\{\|\boldsymbol{L}\boldsymbol{M}^{\dagger}\boldsymbol{N}I_n + \boldsymbol{L}|_{\ker(\boldsymbol{M})}\chi\|^2_{H^l} : \chi \in \ker(\boldsymbol{M})\right\}.
\end{aligned}
$$

By Lemma 2, $\ker(\boldsymbol{M})$ is of finite dimension; therefore, the image of $\boldsymbol{L}|_{\ker(\boldsymbol{M})}$ is finite-dimensional and closed. Thus, $(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}$ exists and is a bounded linear operator (Groetsch, 1977, Cor. 2.1.3). The minimum above is therefore attained for $\chi = -(\boldsymbol{L}|_{\ker(\boldsymbol{M})})^{\dagger}\boldsymbol{L}\boldsymbol{M}^{\dagger}\boldsymbol{N}I_n + \ker(\boldsymbol{L}|_{\ker(\boldsymbol{M})})$ (Groetsch, 1977, p. 41). The uniqueness result then follows from the fact that $\ker(\boldsymbol{L}|_{\ker(\boldsymbol{M})}) = \ker(\boldsymbol{L}) \cap \ker(\boldsymbol{M})$. □

Proof of Theorem 7. Fix $\theta_0 \in \Theta$. Condition (i) together with Lemma 6 implies that

$$
\|\phi_{\boldsymbol{L}}(\theta) - \phi_{\boldsymbol{L}}(\theta_0)\|_{\infty} \le \|\phi_{\boldsymbol{L}}(\theta) - \phi_{\boldsymbol{L}}(\theta_0)\|_{H^m} + \|\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta) - \boldsymbol{d}\phi_{\boldsymbol{L}}(\theta_0)\|_{H^m},
$$

provided $\phi_{\boldsymbol{L}}(\theta)$, $\phi_{\boldsymbol{L}}(\theta_0)$, $\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta)$, and $\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta_0)$ are in H_0^m. We will prove that this is indeed the case and the right-hand side converges to zero as $\theta \to \theta_0$. In this proof, all operator

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

norms and orthogonal projections are understood to be with respect to H_0^m and not with respect to H_0^m. Adjoints and Moore–Penrose inverses with respect to H_0^m will be denoted by the separate notation used in Lemma 8.

STEP 1. $\boldsymbol{\phi_L}(\theta)$ is well defined for every $\theta \in \Theta$ and $\lim_{\theta\to\theta_0} \|\boldsymbol{\phi_L}(\theta) - \boldsymbol{\phi_L}(\theta_0)\|_{H^m} = 0$.

By Lemma 5, conditions (iii), (v), and (vi) imply that $\boldsymbol{\phi_L}(\theta)$ is uniquely defined for any $\theta \in \Theta$, including θ_0.

By the discussion following Definition 1 and condition (iii), the operators $\boldsymbol{M}(\theta)$ and $\boldsymbol{M}(\theta_0)$ are well defined. By the discussion following Definition 7, the operator norm of $\boldsymbol{M}(\theta) - \boldsymbol{M}(\theta_0)$ is bounded above by $\|M(\cdot,\theta) - M(\cdot,\theta_0)\|_\infty$. By condition (iv), $M(z,\theta)$ is jointly continuous and so $\|M(\cdot,\theta) - M(\cdot,\theta_0)\|_\infty$ is continuous at θ_0 (Sundaram, 1996, Thm. 9.14). It follows that $\lim_{\theta\to\theta_0} \|M(\cdot,\theta) - M(\cdot,\theta_0)\|_\infty = 0$ and so $\boldsymbol{M}(\theta)$ converges to $\boldsymbol{M}(\theta_0)$ in the operator norm as $\theta \to \theta_0$. The same argument applied to $\boldsymbol{N}(\theta)$ proves that $\boldsymbol{N}(\theta)$ converges to $\boldsymbol{N}(\theta_0)$ in the operator norm as $\theta \to \theta_0$.

Lemma 2 and condition (vi) imply that $\boldsymbol{M}(\theta)$ is onto for any $\theta \in \Theta$, so the orthogonal projection operator onto the image of $\boldsymbol{M}(\theta)$ is the identity mapping on H_0^m for any $\theta \in \Theta$. Thus, $\boldsymbol{M}(\theta)\boldsymbol{M}(\theta)^\dagger = \boldsymbol{M}(\theta_0)\boldsymbol{M}(\theta_0)^\dagger$ and Theorem 1.6 of Koliha (2001) implies that $\boldsymbol{M}(\theta)^\dagger$ converges in the operator norm to $\boldsymbol{M}(\theta_0)^\dagger$ as $\theta \to \theta_0$.

Next, for $\theta \in \Theta$, let $\boldsymbol{Q}(\theta)$ be the orthogonal projection onto $\ker(\boldsymbol{M}(\theta))$. By verifying the four conditions that determine the Moore–Penrose inverse (Groetsch, 1977, p. 48), it is easily seen that $(\boldsymbol{L}|_{\ker(\boldsymbol{M}(\theta))})^\dagger = (\boldsymbol{LQ}(\theta))^\dagger$. By Lemma 2 and condition (vi), $\boldsymbol{LQ}(\theta)$ is an operator of finite rank. Condition (v) now implies that the rank of $\boldsymbol{LQ}(\theta)$ is equal to $\dim(\ker(\boldsymbol{M}(\theta)))$. Since $\boldsymbol{Q}(\theta) = \boldsymbol{I} - \boldsymbol{M}(\theta)^\dagger\boldsymbol{M}(\theta)$ converges in the operator norm to $\boldsymbol{Q}(\theta_0) = \boldsymbol{I} - \boldsymbol{M}(\theta_0)^\dagger\boldsymbol{M}(\theta_0)$ as $\theta \to \theta_0$, $\boldsymbol{LQ}(\theta)$ converges to $\boldsymbol{LQ}(\theta_0)$ in operator norm and the smallest nonzero singular value of $\boldsymbol{LQ}(\theta)$ also converges to that of $\boldsymbol{LQ}(\theta_0)$ (Gohberg et al., 1990, Cor. VI.1.6). It follows that the operator norm of $(\boldsymbol{LQ}(\theta))^\dagger$ remains bounded as $\theta \to \theta_0$. By Theorem 1.6 of Koliha (2001) again, $(\boldsymbol{LQ}(\theta))^\dagger$ converges in operator norm to $(\boldsymbol{LQ}(\theta_0))^\dagger$.

It follows from the above and Lemma 8(iv) that $\boldsymbol{M}(\theta)^- = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M}(\theta))})^\dagger\boldsymbol{L})\boldsymbol{M}(\theta)^\dagger$ converges in operator norm to $\boldsymbol{M}(\theta_0)^- = (\boldsymbol{I} - (\boldsymbol{L}|_{\ker(\boldsymbol{M}(\theta_0))})^\dagger\boldsymbol{L})\boldsymbol{M}(\theta_0)^\dagger$ as $\theta \to \theta_0$. Therefore, $\boldsymbol{\phi_L}(\theta) = \boldsymbol{M}(\theta)^-\boldsymbol{N}(\theta)I_n$ converges to $\boldsymbol{\phi_L}(\theta_0) = \boldsymbol{M}(\theta_0)^-\boldsymbol{N}(\theta_0)I_n$ in H_0^m as $\theta \to \theta_0$.

STEP 2. $\boldsymbol{d\phi_L}(\theta)$ exists and $\lim_{\theta\to\theta_0} \|\boldsymbol{d\phi_L}(\theta) - \boldsymbol{d\phi_L}(\theta_0)\|_{H^m} = 0$.

For $\theta \in \Theta$, $\boldsymbol{dM}(\theta)$ exists by Lemma 7 and condition (iii). It follows that $\boldsymbol{S}_\omega\boldsymbol{M}(\theta)\boldsymbol{S}_{-\omega}$ converges to $\boldsymbol{M}(\theta)$ in the operator norm as $\omega \to 0$. For $\omega \in \mathbb{R}$, $\boldsymbol{S}_\omega\boldsymbol{M}(\theta)\boldsymbol{S}_{-\omega} : \phi \mapsto P(M(ze^{i\omega},\theta)\phi|H_0^m)$ and $M(ze^{i\omega},\theta) \in \mathcal{W}^{m\times m}$ has the same partial indices as $M(z,\theta)$. Therefore, condition (vi) implies that $\boldsymbol{S}_\omega\boldsymbol{M}(\theta)\boldsymbol{S}_{-\omega}$ is onto and by the same argument as in step 1, $(\boldsymbol{S}_\omega\boldsymbol{M}(\theta)\boldsymbol{S}_{-\omega})^-$ converges to $\boldsymbol{M}(\theta)^-$ in the operator norm as $\omega \to 0$. Theorem 2.1 of Koliha (2001) applied to $\boldsymbol{S}_\omega\boldsymbol{M}(\theta)\boldsymbol{S}_{-\omega}$, as $\mathsf{H}_0^m \to \mathsf{H}_0^m$ mappings, gives

$$\boldsymbol{d}(\boldsymbol{M}(\theta)^-) = -\boldsymbol{M}(\theta)^-\boldsymbol{dM}(\theta)\boldsymbol{M}(\theta)^- + (\boldsymbol{I} - \boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))\boldsymbol{dM}(\theta)^\times(\boldsymbol{M}(\theta)^-)^\times\boldsymbol{M}(\theta)^-.$$

By Lemma 7 and condition (iii) again,

$$\boldsymbol{d}(\boldsymbol{N}(\theta)I_n) = \boldsymbol{dN}(\theta)I_n.$$

This implies that, for every $\theta \in \Theta$,

$$\begin{aligned}\boldsymbol{d\phi_L}(\theta) &= \boldsymbol{d}(\boldsymbol{M}(\theta)^-)\boldsymbol{N}(\theta)I_n + \boldsymbol{M}(\theta)^-\boldsymbol{dN}(\theta)I_n \\ &= \big(-\boldsymbol{M}(\theta)^-\boldsymbol{dM}(\theta)\boldsymbol{M}(\theta)^- + (\boldsymbol{I} - \boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))\boldsymbol{dM}(\theta)^\times(\boldsymbol{M}(\theta)^-)^\times\boldsymbol{M}(\theta)^-\big)\boldsymbol{N}(\theta)I_n \\ &\quad + \boldsymbol{M}(\theta)^-\boldsymbol{dN}(\theta)I_n\end{aligned}$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

$$
\begin{aligned}
&= -\boldsymbol{M}(\theta)^{-}\boldsymbol{dM}(\theta)\boldsymbol{M}(\theta)^{-}\boldsymbol{N}(\theta)I_n \\
&\quad + (\boldsymbol{I} - \boldsymbol{M}(\theta)^{-}\boldsymbol{M}(\theta))\boldsymbol{W}(\theta)^{-1}\boldsymbol{dM}(\theta)^{*}(\boldsymbol{M}(\theta)^{-})^{*}\boldsymbol{W}(\theta)\boldsymbol{M}(\theta)^{-}\boldsymbol{N}(\theta)I_n \\
&\quad + \boldsymbol{M}(\theta)^{-}\boldsymbol{dN}(\theta)I_n,
\end{aligned}
$$

where $\boldsymbol{W}(\theta) = \boldsymbol{M}(\theta)^{*}\boldsymbol{M}(\theta) + \boldsymbol{L}^{*}\boldsymbol{L}$. Clearly, $\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta)$ converges to $\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta_0)$ in H_0^m as $\theta \to \theta_0$ if all of the operators that appear in the expression above are continuous in the operator norm at θ_0. The continuity of $\boldsymbol{M}(\theta)$, $\boldsymbol{N}(\theta)$, and $\boldsymbol{M}(\theta)^{-}$ at θ_0 has already been established in step 1. It remains to establish the operator norm continuity of $\boldsymbol{W}(\theta)^{-1}$, $\boldsymbol{dM}(\theta)$, and $\boldsymbol{dN}(\theta)$ at θ_0. The continuity of inversion at invertible operators ensures that $\boldsymbol{W}(\theta)^{-1}$ is continuous in the operator norm at θ_0 (Gohberg et al., 2003a, Cor. II.8.2). The operator norm of $\boldsymbol{dM}(\theta) - \boldsymbol{dM}(\theta_0)$ is bounded above by $\sup_{z\in\mathbb{T}} \left\| \mathrm{i}z\frac{d}{dz}M(z,\theta) - \mathrm{i}z\frac{d}{dz}M(z,\theta_0) \right\|_{\mathbb{C}^{m\times m}} = \sup_{z\in\mathbb{T}} \left\| \frac{d}{dz}M(z,\theta) - \frac{d}{dz}M(z,\theta_0) \right\|_{\mathbb{C}^{m\times m}}$, which converges to zero as $\theta \to \theta_0$ by the joint continuity of $\frac{d}{dz}M(z,\theta)$ (condition (iv)). A similar argument yields the continuity of $\boldsymbol{dN}(\theta)$. □

Proof of Theorem 8. For $\phi : \Theta \to H^m$, $x \in \mathbb{R}^d$, and $\epsilon \neq 0$, define

$$
\nabla_{\epsilon,x}\phi(\theta) = \frac{\phi(\theta + \epsilon x) - \phi(\theta)}{\epsilon}.
$$

The claim of the theorem is that for every $\theta \in \Theta$, there exists a symmetric multilinear mapping $D_\theta^p\phi_{\boldsymbol{L}}(\theta) : \prod_{i=1}^p \mathbb{R}^d \to H_0^m$ such that for every $\{x_1, \ldots, x_p\} \subset \mathbb{R}^d$,

$$
\lim_{(\epsilon_1,\ldots,\epsilon_p)\to 0} \left\| \nabla_{\epsilon_p,x_p} \cdots \nabla_{\epsilon_1,x_1}\phi_{\boldsymbol{L}}(\theta) - D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p) \right\|_\infty = 0
$$

and $D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p)$ is continuous in μ-essential supremum norm with respect to θ. As in the proof of Theorem 7, we will prove the existence of $\nabla_{\epsilon_p,x_p} \cdots \nabla_{\epsilon_1,x_1}\phi_{\boldsymbol{L}}(\theta)$ and $\boldsymbol{d}\left(\nabla_{\epsilon_p,x_p} \cdots \nabla_{\epsilon_1,x_1}\phi_{\boldsymbol{L}}(\theta)\right)$ and their convergence as elements of H^m to $D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p)$ and $\boldsymbol{d}(D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p))$, respectively, as $(\epsilon_1,\ldots,\epsilon_p) \to 0$. The continuity in H^m of $D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p)$ and $\boldsymbol{d}(D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p))$ with respect to θ then proves the continuity of $D_\theta^p\phi_{\boldsymbol{L}}(\theta)(x_1,\ldots,x_p)$ in μ-essential supremum norm with respect to θ by Lemma 6.

STEP 1. The result holds for $p = 1$.
For $x \in \mathbb{R}^d$ and $\phi \in H_0^m$, define

$$
\boldsymbol{D}_\theta\boldsymbol{M}(\theta)x : \phi \mapsto \boldsymbol{P}((D_\theta M(z,\theta)x)\phi | H_0^m),
$$

where $D_\theta M(z,\theta)$ is the Jacobian of $M(z,\theta)$ with respect to θ. Then, by arguments that are by now familiar,

$$
\begin{aligned}
\left\| \nabla_{\epsilon,x}\boldsymbol{M}(\theta)\phi - (\boldsymbol{D}_\theta\boldsymbol{M}(\theta)x)\phi \right\|_{H^m} &= \left\| \boldsymbol{P}((\nabla_{\epsilon,x}M(z,\theta) - D_\theta M(z,\theta)x)\phi | H_0^m) \right\|_{H^m} \\
&\leq \left\| (\nabla_{\epsilon,x}M(z,\theta) - D_\theta M(z,\theta)x)\phi \right\|_{H^m} \\
&\leq \left\| \nabla_{\epsilon,x}M(z,\theta) - D_\theta M(z,\theta)x \right\|_\infty \|\phi\|_{H^m} \\
&\leq \left\| \frac{1}{\epsilon}\int_0^\epsilon (D_\theta M(z,\theta+\rho x)x - D_\theta M(z,\theta)x)\, d\rho \right\|_\infty \|\phi\|_{H^m} \\
&\leq \sup_{0\leq\rho\leq\epsilon} \|D_\theta M(z,\theta+\rho x)x - D_\theta M(z,\theta)x\|_\infty \|\phi\|_{H^m},
\end{aligned}
$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

which converges to zero as $\epsilon \to 0$ by the uniform continuity of $D_\theta M(z,\theta+\rho x)x$ with respect to $(z,\rho) \in \mathbb{T} \times [0,\bar{\epsilon}]$, where $\bar{\epsilon}$ is chosen so that the segment between θ and $\theta+\bar{\epsilon}x$ is inside Θ. Thus, the mapping $\phi \mapsto (\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\phi$ is linear and bounded on H_0^m (Gohberg et al., 2003a, Cor. XII.4.4). By the same arguments, the mapping $\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x : \phi \mapsto \boldsymbol{P}((D_\theta N(z,\theta)x)\phi | H_0^m)$ is a bounded linear operator from H_0^n to H_0^m as well as the operator limit of $\boldsymbol{\nabla}_{\epsilon,x}\boldsymbol{N}(\theta)$.

Since $\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x$ exists and, by the same arguments as used above, $\boldsymbol{M}(\theta+\epsilon x)^-$ converges to $\boldsymbol{M}(\theta)^-$ as $\epsilon \to 0$, Theorem 2.1 of Koliha (2001), applied to $\boldsymbol{M}(\theta+\epsilon x)$ as $\mathsf{H}_0^m \to \mathsf{H}_0^m$ mappings, implies that $\boldsymbol{\nabla}_{\epsilon,x}(\boldsymbol{M}(\theta)^-)$ converges in operator norm to

$$\begin{aligned}\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x &= -\boldsymbol{M}(\theta)^-(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times(\boldsymbol{M}(\theta)^-)^\times \boldsymbol{M}(\theta)^- \\ &= -\boldsymbol{M}(\theta)^-(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))\boldsymbol{W}(\theta)^{-1}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^*(\boldsymbol{M}(\theta)^-)^*\boldsymbol{W}(\theta)\boldsymbol{M}(\theta)^-,\end{aligned}$$

as $\epsilon \to 0$, where $\boldsymbol{W}(\theta) = \boldsymbol{M}(\theta)^*\boldsymbol{M}(\theta) + \boldsymbol{L}^*\boldsymbol{L}$. It follows that the mapping $\phi \mapsto (\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x)\phi$ is a bounded linear operator on H_0^m. Thus, $\boldsymbol{\nabla}_{\epsilon,x}\boldsymbol{\phi}_{\boldsymbol{L}}(\theta)$ converges in H^m to

$$\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x = (\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x)\boldsymbol{N}(\theta)I_n + \boldsymbol{M}(\theta)^-(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)I_n.$$

Next, following the same techniques as used above, the joint continuity of $\frac{d}{dz}(D_\theta M(z,\theta)x) = D_\theta\left(\frac{d}{dz}M(z,\theta)\right)x$ implies that the operator $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)$ exists as a cross operator derivative, is equal to the cross derivative $\boldsymbol{D}_\theta(\boldsymbol{dM}(\theta)x)$, is given by $\phi \mapsto \boldsymbol{P}\left(\mathrm{i}z\frac{d}{dz}(D_\theta M(z,\theta)x)\phi \middle| H_0^m\right)$, and is a bounded linear operator on H_0^m. The same is true of $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)$. This implies that $\boldsymbol{d}(\boldsymbol{\nabla}_{\epsilon,x}\boldsymbol{\phi}_{\boldsymbol{L}}(\theta)) = \boldsymbol{\nabla}_{\epsilon,x}\boldsymbol{d}\boldsymbol{\phi}_{\boldsymbol{L}}(\theta)$ converges in H^m to $\boldsymbol{D}_\theta(\boldsymbol{d}\boldsymbol{\phi}_{\boldsymbol{L}}(\theta))x = \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x)$, given by

$$\begin{aligned}\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x) &= \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x)\boldsymbol{N}(\theta)I_n + (\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x)\boldsymbol{dN}(\theta)I_n \\ &\quad + \boldsymbol{d}(\boldsymbol{M}(\theta)^-)(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)I_n + \boldsymbol{M}(\theta)^-\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)I_n \\ &= \big\{-\boldsymbol{d}(\boldsymbol{M}(\theta)^-)(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\boldsymbol{M}(\theta)^- - \boldsymbol{M}(\theta)^-\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\boldsymbol{M}(\theta)^- \\ &\quad - \boldsymbol{M}(\theta)^-(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)\boldsymbol{d}(\boldsymbol{M}(\theta)^-) + \\ &\quad + (\boldsymbol{I}-\boldsymbol{d}(\boldsymbol{M}(\theta)^-)\boldsymbol{M}(\theta))(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times(\boldsymbol{M}(\theta)^-)^\times \boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^- d\boldsymbol{M}(\theta))(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times(\boldsymbol{M}(\theta)^-)^\times \boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times(\boldsymbol{M}(\theta)^-)^\times \boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times\boldsymbol{d}(\boldsymbol{M}(\theta)^-)^\times \boldsymbol{M}(\theta)^- \\ &\quad + (\boldsymbol{I}-\boldsymbol{M}(\theta)^-\boldsymbol{M}(\theta))(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)^\times(\boldsymbol{M}(\theta)^-)^\times\boldsymbol{d}(\boldsymbol{M}(\theta)^-)\big\}\boldsymbol{N}(\theta)I_n + \\ &\quad + (\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x)\boldsymbol{dN}(\theta)I_n + \boldsymbol{d}(\boldsymbol{M}(\theta)^-)(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)I_n \\ &\quad + \boldsymbol{M}(\theta)^-\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)I_n.\end{aligned}$$

Given the expressions for $\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x$, $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)x)$, and $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)x)$, as well as results obtained in the proof of Theorem 7, it is now clear that all of the operators that appear in the expressions for $\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x$ and $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x)$ are continuous in operator norm with respect to θ. It follows that $\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x$ and $\boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{\phi}_{\boldsymbol{L}}(\theta)x)$ are continuous in H^m-norm with

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

respect to θ. By Lemma 6 then, $\boldsymbol{D}_\theta \phi_{\boldsymbol{L}}(\theta)x$ is continuous in μ-essential supremum norm with respect to θ.

STEP 2. The result holds for all $p > 1$.

For $\{x_1, \ldots, x_p\} \subset \mathbb{R}^d$, $\phi \in H_0^m$, and $i = 1, \ldots, p$, define

$$\boldsymbol{D}_\theta^i \boldsymbol{M}(\theta)(x_1, \ldots, x_i) : \phi \mapsto \boldsymbol{P}(D_\theta^i M(z, \theta)(x_1, \ldots, x_i)\phi | H_0^m),$$

where $D_\theta^i M(z,\theta)(x_1, \ldots, x_i) = \lim_{\epsilon_i \to 0, \ldots, \epsilon_1 \to 0} \nabla_{\epsilon_i, x_i} \cdots \nabla_{\epsilon_1, x_1} M(z, \theta)$ is the ith order derivative of $M(z, \theta)$ with respect to θ in the directions $x_1, \ldots, x_i$. Just as in step 1, it is easily established that for $\{x_1, \ldots, x_p\} \subset \mathbb{R}^d$ and $i = 1, \ldots, p$, $\boldsymbol{D}_\theta^i \boldsymbol{M}(\theta)(x_1, \ldots, x_i)$ is the operator limit of $\nabla_{\epsilon_i, x_i} \cdots \nabla_{\epsilon_1, x_1} \boldsymbol{M}(\theta)$ as $(\epsilon_1, \ldots, \epsilon_i) \to 0$; therefore, it is linear and bounded on H_0^m. The same result holds for $\boldsymbol{D}_\theta^i \boldsymbol{N}(\theta)(x_1, \ldots, x_i)$.

The expression of $\boldsymbol{D}_\theta \boldsymbol{M}(\theta)^- x$ in step 1 consists of further differentiable operators. Therefore, for $\{x_1, \ldots, x_p\} \subset \mathbb{R}^d$ and $i = 1, \ldots, p$, the operator limit of $\nabla_{\epsilon_i, x_i} \cdots \nabla_{\epsilon_1, x_1} \boldsymbol{M}(\theta)^-$ exists, call it $\boldsymbol{D}_\theta^i \boldsymbol{M}(\theta)^-(x_1, \ldots, x_i)$, and is a finite sum of composites of the mappings

$$\begin{aligned}&\boldsymbol{M}(\theta), \boldsymbol{M}(\theta)^-, \boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_1), \ldots, \boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_p), \boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_1, x_2), \ldots, \boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_{i-1}, x_i),\\ &\quad \boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_1, x_2, x_3), \ldots, \boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_{i-2}, x_{i-1}, x_i), \ldots, \boldsymbol{D}_\theta^i \boldsymbol{M}(\theta)(x_1, \ldots, x_i),\end{aligned}$$

and their H_0^m adjoints. As all of these mappings are continuous in the operator norm with respect to θ, it follows that $\boldsymbol{D}_\theta^i \phi_{\boldsymbol{L}}(z, \theta)(x_1, \ldots, x_i)$ is continuous in H^m with respect to θ.

Next, the joint continuity of $\frac{d}{dz} D_\theta^i M(z, \theta)(x_1, \ldots, x_i) = D_\theta^i \left(\frac{d}{dz} M(z, \theta) \right)(x_1, \ldots, x_i)$, for $i = 1, \ldots, p$, implies the existence of the operator derivative

$$\boldsymbol{d}(\boldsymbol{D}_\theta^i \boldsymbol{M}(\theta)(x_1, \ldots, x_i)) : \phi \mapsto \boldsymbol{P}\left(\mathrm{i}z \frac{d}{dz} D_\theta^i M(z, \theta)(x_1, \ldots, x_i)\phi \middle| H_0^m \right),$$

for $i = 1, \ldots, p$, a similar result also holding for $\boldsymbol{d}(\boldsymbol{D}_\theta^i \boldsymbol{N}(\theta)(x_1, \ldots, x_i))$. This implies that

$$\boldsymbol{d}(\nabla_{\epsilon_p, x_p} \cdots \nabla_{\epsilon_1, x_1} \phi_{\boldsymbol{L}}(\theta)) = \nabla_{\epsilon_p, x_p} \cdots \nabla_{\epsilon_1, x_1} \boldsymbol{d}\phi_{\boldsymbol{L}}(\theta)$$

converges in H^m to

$$\boldsymbol{d}(\boldsymbol{D}_\theta^p \phi_{\boldsymbol{L}}(\theta)(x_1, \ldots, x_p)) = \boldsymbol{D}_\theta^p (\boldsymbol{d}\phi_{\boldsymbol{L}}(\theta))(x_1, \ldots, x_p).$$

In particular, $\boldsymbol{d}(\boldsymbol{D}_\theta^p \phi_{\boldsymbol{L}}(z, \theta_0)(x_1, \ldots, x_p))$ is a finite sum of composites of the mappings

$$\begin{aligned}&\boldsymbol{M}(\theta), \boldsymbol{M}(\theta)^-, \boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_1), \ldots, \boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_p), \boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_1, x_2), \ldots, \boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_{p-1}, x_p),\\ &\quad \boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_1, x_2, x_3), \ldots, \boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_{p-2}, x_{p-1}, x_p), \ldots, \boldsymbol{D}_\theta^p \boldsymbol{M}(\theta)(x_1, \ldots, x_p),\\ &\boldsymbol{d}\boldsymbol{M}(\theta), \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_1)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{M}(\theta)(x_p)), \boldsymbol{d}(\boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_1, x_2)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^2 \boldsymbol{M}(\theta)(x_{p-1}, x_p)),\\ &\quad \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_1, x_2, x_3)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_{p-2}, x_{p-1}, x_p)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^p \boldsymbol{M}(\theta)(x_1, \ldots, x_p)),\\ &\quad\quad \boldsymbol{N}(\theta), \boldsymbol{D}_\theta \boldsymbol{N}(\theta)(x_1), \ldots, \boldsymbol{D}_\theta \boldsymbol{N}(\theta)(x_p), \boldsymbol{D}_\theta^2 \boldsymbol{N}(\theta)(x_1, x_2), \ldots, \boldsymbol{D}_\theta^2 \boldsymbol{N}(\theta)(x_{p-1}, x_p),\\ &\quad \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{N}(\theta)(x_1, x_2, x_3)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_{p-2}, x_{p-1}, x_i)), \ldots, \boldsymbol{D}_\theta^p \boldsymbol{N}(\theta)(x_1, \ldots, x_p),\\ &\boldsymbol{d}\boldsymbol{N}(\theta), \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)(x_1)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta \boldsymbol{N}(\theta)(x_p)), \boldsymbol{d}(\boldsymbol{D}_\theta^2 \boldsymbol{N}(\theta)(x_1, x_2)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^2 \boldsymbol{N}(\theta)(x_{p-1}, x_p)),\\ &\quad \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{N}(\theta)(x_1, x_2, x_3)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^3 \boldsymbol{M}(\theta)(x_{p-2}, x_{p-1}, x_p)), \ldots, \boldsymbol{d}(\boldsymbol{D}_\theta^p \boldsymbol{N}(\theta)(x_1, \ldots, x_p))\end{aligned}$$

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

and their H_0^m adjoints applied to I_n. All of these operators are bounded on H_0^m and continuous in the operator norm at θ_0 by the joint continuity of $D_\theta^i M(z,\theta)$, $\frac{d}{dz}D_\theta^i M(z,\theta)$, $D_\theta^i N(z,\theta)$, and $\frac{d}{dz}D_\theta^i N(z,\theta)$ for $i = 0,\ldots,p$. Thus, $\boldsymbol{d}(\boldsymbol{D}_\theta^p \boldsymbol{\phi}_{\boldsymbol{L}}(\theta))$ is continuous at θ_0 in H^m and the result follows by Lemma 6. □

REFERENCES

Al-Sadoon, M. M. (2018). The linear systems approach to linear rational expectations models. *Econometric Theory*, 34(3), 628–658.

Al-Sadoon, M. M. (2019). Sims vs. Onatski. https://majidalsadoon.wordpress.com/2019/10/25/sims-vs-onatski

Al-Sadoon, M. M. (2020). *Regularized solutions to linear rational expectations models*. Preprint, arXiv:2009.05875.

Al-Sadoon, M. M., & Zwiernik, P. (2019). *The identification problem for linear rational expectations models*. Preprint, arXiv:1908.09617.

Anderson, B. (1985). Continuity of the spectral factorization operation. *Matemática Aplicada e Computacional*, 4(2), 139–156.

Arnold, V. I. (1973). *Ordinary differential equations*. MIT Press.

Baggio, G., & Ferrante, A. (2016). On the factorization of rational discrete-time spectral densities. *IEEE Transactions on Automatic Control*, 61(4), 969–981.

Bianchi, F., & Nicolò, G. (2021). A generalized approach to indeterminacy in linear rational expectations models. *Quantitative Economics*, 12(3), 843–868.

Bingham, N. (2012a). Multivariate prediction and matrix Szegö theory. *Probability Surveys*, 9, 325–339.

Bingham, N. (2012b). Szegö's theorem and its probabilistic descendants. *Probability Surveys*, 9, 287–324.

Brockwell, P. J., & Davis, R. A. (1991). *Time series: Theory and methods* (2nd ed.). Springer.

Broze, L., Gourieroux, C., & Szafarz, A. (1985). Solutions of linear rational expectations models. *Econometric Theory*, 1(3), 341–368.

Broze, L., Gouriéroux, C., & Szafarz, A. (1995). Solutions of multivariate rational expectations models. *Econometric Theory*, 11, 229–257.

Cagan, P. (1956). The monetary dynamics of hyperinflation. In M. Friedman (Ed.), *Studies in the quantity theory of money* (pp. 25–117). University of Chicago Press.

Caines, P. E. (1988). *Linear stochastic systems*. Society for Industrial and Applied Mathematics (reprinted 2018).

Callon, G., & Groetsch, C. (1987). The method of weighting and approximation of restricted pseudosolutions. *Journal of Approximation Theory*, 51(1), 11–18.

Canova, F. (2011). *Methods for applied macroeconomic research*. Princeton University Press.

Chahrour, R., & Jurado, K. (2021). Recoverability and expectations-driven fluctuations. *The Review of Economic Studies*, 89(1), 214–239.

Christiano, L. J., & Vigfusson, R. J. (2003). Maximum likelihood in the frequency domain: The importance of time-to-plan. *Journal of Monetary Economics*, 50(4), 789–815.

Clancey, K. F., & Gohberg, I. (1981). *Factorization of matrix functions and singular integral operators*. Operator Theory: Advances and Applications (Vol. 3). Birkhäuser.

Cramér, H. (1940). On the theory of stationary random processes. *Annals of Mathematics*, 41, 215–230.

Cramér, H. (1942). On harmonic analysis in certain functional spaces. *Arkiv för Matematik, Astronomi och Fysik*, 28B(12), Article 17.

Deistler, M., & Pötscher, B. M. (1984). The behaviour of the likelihood function for ARMA models. *Advances in Applied Probability*, 16(4), 843–866.

DeJong, D., & Dave, C. (2011). *Structural macroeconometrics* (2nd ed.). Princeton University Press.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Ephremidze, L., Shargorodsky, E., & Spitkovsky, I. (2020). Quantitative results on continuity of the spectral factorization mapping. *Journal of the London Mathematical Society*, 101(1), 60–81.

Fanelli, L. (2012). Determinacy, indeterminacy and dynamic misspecification in linear rational expectations models. *Journal of Econometrics*, 170(1), 153–163.

Farmer, R. E., Khramov, V., & Nicolò, G. (2015). Solving and estimating indeterminate DSGE models. *Journal of Economic Dynamics and Control*, 54(C), 17–36.

Farmer, R. E. A. (1999). *Macroeconomics of self-fulfilling prophecies* (2nd ed.). MIT University Press.

Funovits, B. (2017). The full set of solutions of linear rational expectations models. *Economics Letters*, 161, 47–51.

Funovits, B. (2020). *The dimension of the set of causal solutions of linear multivariate rational expectations models*. Preprint, arXiv:2002.04369.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, 129(4), 1661–1710.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: An introduction to the new Keynesian framework and its applications* (2nd ed.). Princeton University Press.

Gohberg, I., Goldberg, S., & Kaashoek, M. A. (1990). *Classes of linear operators* (vol. 1). Birkhäuser Verlag.

Gohberg, I., Goldberg, S., & Kaashoek, M. A. (1993). *Classes of linear operators* (vol. 2). Birkhäuser Verlag.

Gohberg, I., Goldberg, S., & Kaashoek, M. A. (2003a). *Basic classes of linear operators*. Berkäuser Verlag.

Gohberg, I., Kaashoek, M. A., & Spitkovsky, I. M. (2003b). An overview of matrix factorization theory and operator applications. In I. Gohberg, N. Manojlovic, & A. F. dos Santos (Eds.), *Factorization and integrable systems: Summer school in Faro, Portugal, September 2000*. Operator Theory: Advances and Applications (vol. 141), chapter 1 (pp. 1–102). Springer Basel AG.

Gohberg, I. C., & Fel'dman, I. A. (1974). *Convolution equations and projection methods for their solution*. Translations of Mathematical Monographs (vol. 41). American Mathematical Society.

Green, M., & Anderson, B. D. (1987). On the continuity of the Wiener–Hopf factorization operation. *The Journal of the Australian Mathematical Society. Series B. Applied Mathematics*, 28(4), 443–461.

Groetsch, C. (1977). *Generalized inverses of linear operators: Representation and approximation*. Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker, Inc.

Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13, 49–52.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability*, 10(1), 130–145.

Hannan, E. J., & Deistler, M. (1988). *The statistical theory of linear systems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (reprinted 2012).

Hansen, L. P., & Sargent, T. J. (1980). Formulating and estimating dynamic linear rational expectations models. *Journal of Economic Dynamics and Control*, 2, 7–46.

Herbst, E. P., & Schorfheide, F. (2016). *Bayesian estimation of DSGE models*. Princeton University Press.

Hodrick, R. J., & Prescott, E. C. (1997). Postwar US business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.

Horn, R. A., & Johnson, C. R. (1985). *Matrix analysis*. Cambridge University Press.

Kailath, T. (1980). *Linear systems*. Prentice Hall.

Kalman, R., Arbib, M., & Falb, P. (1969). *Topics in mathematical system theory*. International Series in Pure and Applied Mathematics. McGraw-Hill.

Keynes, J. M. (1936). *The general theory of employment, interest, and money*. Palgrave Macmillan (reprinted 2018).

Knight, F. H. (1921). *Risk, uncertainty and profit*. Hart, Schaffner & Marx Prize Essays. Houghton Mifflin.

Kociecki, A., & Kolasa, M. (2018). Global identification of linearized DSGE models. *Quantitative Economics*, 9(3), 1243–1263.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Kociecki, A., & Kolasa, M. (2023). A solution to the global identification problem in DSGE models. *Journal of Econometrics*, 236(2), Article 105477.

Koliha, J. J. (2001). Continuity and differentiability of the Moore–Penrose inverse in C*-algebras. *Mathematica Scandinavica*, 88(1), 154–160.

Kolmogorov, A. N. (1939). Sur l'interpolation et extrapolation des suites stationnaires. *Comptes rendus de l'Académie des Sciences Paris*, 208, 2043–2045.

Kolmogorov, A. N. (1941a). Interpolation and extrapolation of stationary random sequences. *Izvestiya Akademii Nauk SSSR Seriya Matematicheskaya*, 5, 3–14.

Kolmogorov, A. N. (1941b). Stationary sequences in hilbert space. *Moscow University Mathematics Bulletin*, 2(6), 1–40.

Komunjer, I., & Ng, S. (2011). Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, 79(6), 1995–2032.

Lindquist, A., & Picci, G. (2015). *Linear stochastic systems: A geometric approach to modeling, estimation, and identification*. Series in Contemporary Mathematics (vol. 1). Springer-Verlag.

Locker, J., & Prenter, P. (1980). Regularization with differential operators. I. General theory. *Journal of Mathematical Analysis and Applications*, 74(2), 504–529.

Lubik, T. A., & Schorfheide, F. (2003). Computing sunspot equilibria in linear rational expectations models. *Journal of Economic Dynamics and Control*, 28(2), 273–285.

McCallum, B. T. (1983). On non-uniqueness in rational expectations models: An attempt at perspective. *Journal of Monetary Economics*, 11(2), 139–168.

Meyer-Gohde, A., & Tzaawa-Krenzler, M. (2023). *Sticky information and the Taylor rule* [Technical report]. IMFS Working Paper Series.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29(3), 315–335.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics*(volume 4, chapter 36, pp. 2111–2245). Elsevier.

Nikolski, N. K. (2002). *Operators, functions, and systems: An easy reading: Volume 1: Hardy, Hankel, and Toeplitz*. Mathematical Surveys and Monographs (vol. 92). American Mathematical Society.

Onatski, A. (2006). Winding number criterion for existence and uniqueness of equilibrium in linear rational expectations models. *Journal of Economic Dynamics and Control*, 30(2), 323–345.

Pesaran, M. H. (1987). *The limits to rational expectations*. Basil Blackwell Inc.

Pötscher, B. M., & Prucha, I. R. (1997). *Dynamic nonlinear econometric models: Asymptotic theory*. Springer.

Pourahmadi, M. (2001). *Foundations of time series analysis and prediction theory*. John Wiley & Sons.

Qu, Z., & Tkachenko, D. (2012). Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics*, 3(1), 95–132.

Qu, Z., & Tkachenko, D. (2017). Global identification in DSGE models allowing for indeterminacy. *The Review of Economic Studies*, 84(3), 1306–1345.

Rogosin, S., & Mishuris, G. (2016). Constructive methods for factorization of matrix-functions. *IMA Journal of Applied Mathematics*, 81(2), 365–391.

Rotnitzky, A., Cox, D. R., Bottai, M., & Robins, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, 6(2), 243–284.

Rozanov, Y. A. (1967). *Stationary random processes*. Holden-Day.

Rudin, W. (1986). *Real and complex analysis* (3rd ed.). McGraw-Hill.

Sala, L. (2015). DSGE models in the frequency domains. *Journal of Applied Econometrics*, 30(2), 219–240.

Sargent, T. J. (1979). *Macroeconomic theory*. Academic Press.

Sims, C. A. (2002). Solving linear rational expectations models. *Computational Economics*, 20(1), 1–20.

Sims, C. A. (2007). *On the genericity of the winding number criterion for linear rational expectations models* [Mimeo].

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press

Smets, F., & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, 97(3), 586–606.

Sontag, E. D. (1998). *Mathematical control theory: Deterministic finite dimensional systems* (2nd ed.). Texts in Applied Mathematics. Springer Verlag. www.math.rutgers.edu/~sontag/

Sorge, M. (2019). Arbitrary initial conditions and the dimension of indeterminacy in linear rational expectations models. *Decisions in Economics and Finance*, 43, 363–372.

Sundaram, R. K. (1996). *A first course in optimization theory*. Cambridge University Press.

Tan, F. (2019). A frequency-domain approach to dynamic macroeconomic models. *Macroeconomic Dynamics*, 25(6), 1381–1411.

Tan, F., & Walker, T. B. (2015). Solving generalized multivariate linear rational expectations models. *Journal of Economic Dynamics and Control*, 60, 95–111.

Taylor, J. B. (1977). Conditions for unique solutions in stochastic macroeconomic models with rational expectations. *Econometrica*, 45(6), 1377–1385.

Whiteman, C. H. (1983). *Linear rational expectations models: A user's guide*. University of Minnesota Press.

Wiener, N., & Hopf, E. (1931). Über eine klasse singulärer integralgleichungen. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin*, 31, 696–706.

Wold, H. (1938). *A study in the analysis of stationary time series*. Almqvist and Wiksell.

https://doi.org/10.1017/S026646662400029X Published online by Cambridge University Press