Determinacy and Large-Scale Solutions in the Sequence Space

Adrien Auclert* Evan Majic† Matthew Rognlie‡ Ludwig Straub§

December 2025

Abstract

We show that the sequence-space Jacobians of stationary models have a special "quasi-Toeplitz" form. This result implies a simple test for existence and uniqueness of solutions, helps significantly reduce truncation error, and allows the solution of very large sequence-space systems. We apply these insights to a heterogeneous-agent New Keynesian model, showing how to identify thresholds for determinacy and also how to solve the model using fast iterations rather than direct operations on truncated matrices. We leverage these results to solve a 190-country version of the model with a realistic trade network almost instantly, in spite of its state space having dimension of almost 1 million.

*Stanford University, CEPR, and NBER. Email: aauclert@stanford.edu.

†Northwestern University. Email: majic@u.northwestern.edu.

‡Northwestern University and NBER. Email: matthew.rognlie@northwestern.edu.

§Harvard University, CEPR, and NBER. Email: ludwigstraub@fas.harvard.edu.

An earlier version of this work, which this paper supersedes, circulated under the title "Determinacy and Existence in the Sequence Space". We thank Fernando Alvarez, Anmol Bhandari, Lars Peter Hansen, Greg Kaplan, and participants at the 2023 SED meeting and the 2024 Economic Dynamics, Uncertainty, and Computation conference for useful comments. This research has been supported by National Science Foundation grants SES-2042691 and SES-2343935, and was also conducted in part when Matthew Rognlie was visiting the Federal Reserve Bank of Minneapolis. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

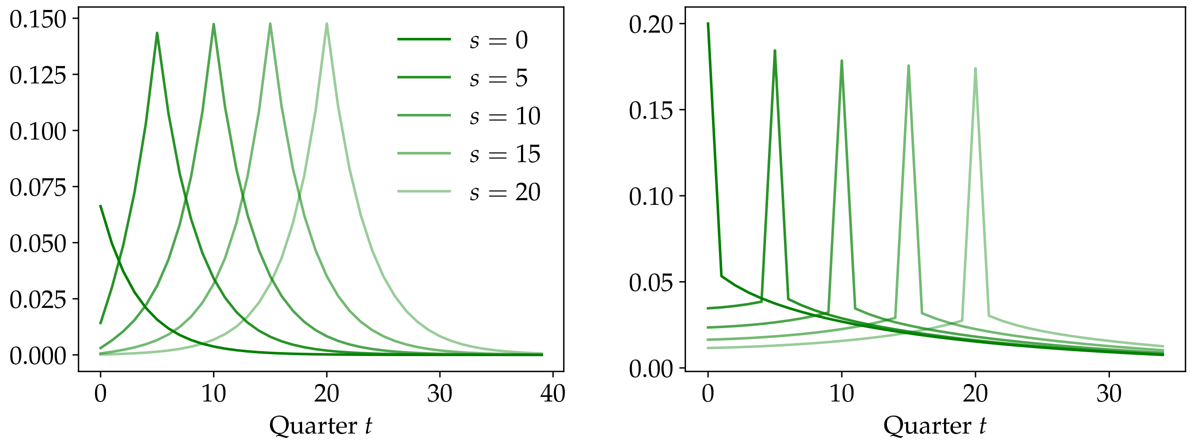
1 Introduction

In macroeconomics, it has become increasingly common for researchers to solve their models in the sequence space: the space of perfect-foresight impulse responses. Due to certainty equivalence, to first order these are the same as the impulse responses in a fully stochastic model (Boppart, Krusell and Mitman 2018). We can solve for them by setting up a linear system in the sequence space, where the mappings between sequences are *sequence-space Jacobians*. This approach is particularly useful for heterogeneous-agent models, where the state space can be very large and sequences offer a more parsimonious representation of equilibrium. Using sequence-space Jacobians, researchers can solve rich heterogeneous-agent models at speeds that are otherwise unachievable (Auclert, Bardóczy, Rognlie and Straub 2021).

At the same time, the sequence-space approach currently suffers from some major limitations. First, there does not exist a standard criterion for determinacy and existence of solutions in the sequence space. Second, in principle, we are solving for sequences with infinite length, and sequence-space Jacobians are matrices with dimension $\infty \times \infty$. In practice, these matrices are truncated to some maximum horizon T, but to avoid errors from truncation, this T—and therefore the sequence-space system—sometimes needs to be quite large. Third, a typical model requires simultaneously solving for n unknown sequences, and in some models—such as network or trade models—n can be very large. For these models, it may be prohibitively expensive to solve the $nT \times nT$ sequence-space system.

In this paper, we overcome these limitations by proving and exploiting a structure result for sequence-space Jacobians. We show that under general conditions, these Jacobians are *quasi-Toeplitz operators* on the sequence space, and apply this structure result in three ways. First, we derive a simple "winding number" criterion for existence and uniqueness, building on Onatski (2006). Second, we obtain a much more parsimonious way to store and handle Jacobians, opening the door to more efficient, truncation-free computations. Finally, we provide a useful preconditioner to solve models iteratively, enabling the solution of even enormous models—such as, in our application, a 190-country fiscal policy model, with heterogeneous agents in each country—in a matter of seconds.

Structure result. We start with the space of square-summable sequences, ℓ^2, which contains all impulse responses consistent with finite variance in a stochastic economy. Our Jacobians are then linear operators mapping ℓ^2 to itself. We say an operator $\mathbf{J} = T(\mathbf{j})$ is *Toeplitz* if each diagonal in its matrix representation is constant. A Toeplitz operator is



(a) Jacobian of P_t vs. MC_s in Calvo model (b) Jacobian of C_t vs. income Y_s in HA model

Figure 1: Examples of quasi-Toeplitz sequence-space Jacobians

characterized by the two-sided sequence $\mathbf{j} = \{\ldots, j_{-1}, j_0, j_1, \ldots\}$ that gives the value on each diagonal, $J_{t,s} = j_{t-s}$. A *quasi*-Toeplitz operator is a Toeplitz operator plus a compact operator $\mathbf{E}$: $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$. The matrix entries $E_{t,s}$ of a compact operator go to zero as $t, s \to \infty$, so a quasi-Toeplitz Jacobian looks Toeplitz for shocks that occur far from zero—in other words, shocks that are sufficiently well-anticipated.

Why are quasi-Toeplitz operators so special? The Jacobians of simple macro equations are often Toeplitz, but composing any two macro equations involves the product of two Toeplitz operators, which is quasi-Toeplitz. We provide an economic interpretation for the new term $\mathbf{E}$: it represents the effect of *missing anticipation* for a shock that is not known before $t = 0$. For instance, figure 1(a) plots columns of the Jacobian mapping nominal marginal cost to aggregate prices in the Calvo model. For shocks to marginal cost at high enough s, the columns of this Jacobian are shifted versions of each other, with $J_{t,s}$ depending only on $t - s$, as in a Toeplitz matrix. But for the unexpected shock to marginal cost at $s = 0$, the aggregate price response is smaller, since sticky price-setters do not react prior to date 0.

We show that, as long as the effects of missing anticipation in a model eventually die out, its Jacobians will be quasi-Toeplitz. We use this insight to prove two key structure theorems. First, we show that for any stationary heterogeneous-agent block, sequence-space Jacobians are quasi-Toeplitz. As an example, figure 1(b) plots the Jacobian of aggregate consumption with respect to income in a standard heterogeneous-agent model. Second, and even more generally, we show that as long as a system of expectational linear difference equations satisfies the usual conditions for a stable solution, that solution can be represented with quasi-Toeplitz Jacobians. These theorems—together with the fact that the quasi-Toeplitz class is closed under addition and multiplication—imply that the vast majority of sequence-space Jacobians in economic applications are quasi-Toeplitz. We apply this result in three ways.

Existence and uniqueness of solutions. To solve for equilibrium, we generally must solve for $\mathbf{x}$ in a system of the form $\mathbf{J}\mathbf{x} = \mathbf{y}$, where $\mathbf{J}$ is a sequence-space Jacobian. There are two key questions one can ask about this system: does a solution exist, and are solutions unique? If $\mathbf{J}$ is Toeplitz, then it is well-known that the *winding number* of $j(z) \equiv \sum_{k=-\infty}^{\infty} j_k z^k$, the number of times the graph of $j(z)$ wraps counterclockwise around the origin in the complex plane as z goes counterclockwise around the unit circle, gives an exact answer. A positive winding number implies nonexistence, and a negative winding number implies indeterminacy; there is a unique solution if and only if the winding number is zero. We show that this winding number criterion extends "generically" to quasi-Toeplitz $\mathbf{J}$—meaning that mathematically, it applies for almost every possible $\mathbf{J}$. The non-generic cases where the criterion fails are simple to detect, because they always involve nonexistence of solutions, which can be easily tested. Importantly, these results all extend to the "block" quasi-Toeplitz case, where $\mathbf{x}$ and $\mathbf{y}$ each stack multiple sequences.

For linear difference equations in canonical form, this winding number test is equivalent to standard root-counting tests like Blanchard and Kahn (1980), as previously shown by Onatski (2006). The winding number test is more general, however, and can be applied directly in the sequence space, allowing us to bypass high-dimensional state spaces. We can also reuse information across multiple calculations: for instance, when we use the winding number test to evaluate determinacy in a heterogeneous-agent model, only trivial computations are needed to redo the test for different Taylor rule coefficients ϕ and Phillips curve slopes κ.

Limiting truncation error. In practice, computations with Jacobians are usually done by truncating them to some finite horizon T. For large enough T, the resulting solutions generally converge to the truth—but in practice this requires fine-tuning T, and can also involve impractically large systems in some cases. An alternative is to bypass large matrix operations by directly exploiting quasi-Toeplitz structure. We provide two complementary approaches. First, we can replace direct solutions with much more efficient iterative solutions by using the Toeplitz part of the inverse—which is cheap to calculate—as a "preconditioner" for iterative methods, such as Krylov subspace methods. Second, we can avoid working with large, truncated $T \times T$ matrices by splitting them into Toeplitz and compact parts, each of which can be handled efficiently: the Toeplitz part with standard techniques, and the compact part via a low-rank approximation.

Solving large-scale systems. Finally, we extend our results to the practical $n > 1$ case, in which multiple aggregate sequences are solved for simultaneously. To illustrate how our computational insights can be useful in models with very large n, we solve for the propagation of fiscal policy in a large multi-country model, where 190 countries consume each others' output according to a trade matrix we take from the data. Although this model is far too large to be solved directly as a sequence-space system, using an iterative approach we obtain impulse responses in just a few seconds on a laptop.

Related literature. Our winding number test is closely connected to, and builds upon, the test introduced by Onatski (2006) for block Toeplitz systems. Our most important contribution is to extend the test to block *quasi*-Toeplitz systems, making it readily applicable to a larger space of models, including heterogeneous-agent models for which other tests are impractical. We also address the Sims (2007) critique of genericity in Onatski (2006), showing that while non-generic economic models do exist, these models suffer from both nonexistence and indeterminacy—and checking for nonexistence is relatively simple. Finally, we build upon Onatski (2006) by showing that when a solution exists, it is itself given by a block quasi-Toeplitz mapping, and by providing ways to efficiently reuse computations when evaluating the winding number.[1]

Although the literature on sequence-space solution methods in economics is rapidly growing, prior to this paper there has been little direct use of Toeplitz or quasi-Toeplitz structure. An early working paper version of Auclert et al. (2021), superseded by this paper, showed that heterogeneous-agent blocks have asymptotically Toeplitz Jacobians and derived a winding number test.[2] Wolf (2025) exploits Toeplitz structure in analytical HA models to prove that the implied consumption Jacobians are invertible. Auclert, Cai, Rognlie and Straub (2024b) use the discounted Toeplitz part of Jacobians (i.e. $j(\beta) = \sum_{k=-\infty}^{\infty} \beta^k j_k$) to characterize the steady state of optimal fiscal policy with commitment. In the mathematical literature, by contrast, there is both a longstanding body of work on Toeplitz operators (Böttcher and Silbermann 2012) and also a more recent practical literature on using quasi-Toeplitz structure for computation (Bini, Massei and Robol 2019).

Our paper relates to an older literature on rational expectations models in the time

[1] In related work, Hagedorn (2023) provides a different local determinacy criterion for incomplete markets models, which relies on dimension reduction.

[2] See Auclert, Bardóczy, Rognlie and Straub (2019), who called this structure—which is implied by, but weaker than, quasi-Toeplitz—"asymptotic time invariance".

and frequency domain, including work by Hansen and Sargent (1980a), Hansen and Sargent (1980b), and Whiteman (1985). This literature studies equilibrium by using the z-transforms of impulse responses. The mappings between these z-transforms can be interpreted as the action of quasi-Toeplitz operators, but this interpretation has not previously been made explicit. More recent papers in this tradition include Tan and Walker (2015), Huo and Takayama (2024), and Meyer-Gohde (2024). Relative to this literature, our paper focuses less on analytical solutions, and more on models that require numerical computation, especially heterogeneous-agent models.

Finally, our use of a Krylov subspace method (GMRES) to iteratively solve large-scale systems connects us with a small literature using related methods in macroeconomics, including Mrkaic (2002) and Gilli and Pauletto (1998) for iterative computation, and Reiter (2010) and Ahn, Kaplan, Moll, Winberry and Wolf (2018) for model reduction.

Layout. The paper is structured as follows. Section 2 introduces Toeplitz and quasi-Toeplitz operators and derives our structure theorems. Section 3 derives and applies the winding number test for quasi-Toeplitz operators. Section 4 shows how quasi-Toeplitz structure can be exploited to avoid large matrix computations, and section 5 applies these methods to solve a large multi-country model. Section 6 concludes.

2 Quasi-Toeplitz sequence-space Jacobians

2.1 Sequence-space Jacobians as operators on ℓ^2

As Auclert, Rognlie and Straub (2024a) show, in some models it is possible to derive an aggregate consumption function $\mathcal{C}_t(\{Z_s\})$, which maps the path of aggregate after-tax income $\{Z_s\}_{s=0}^{\infty}$, starting at date 0, to the resulting path of aggregate consumption $\{C_t\}_{t=0}^{\infty}$, assuming a steady state prior to date 0. This function gives the consumption effect of a perfect-foresight "MIT shock" to income—and as Boppart et al. (2018) point out, to first order, this is the same as the impulse response of C to Z in a fully stochastic rational expectations model.[3]

The derivative of $\mathcal{C}$, the *sequence-space Jacobian* $\mathbf{M}$, maps first-order income impulses $d\mathbf{Z} \equiv \{dZ_s\}_{s=0}^{\infty}$ around the steady state to consumption impulses $d\mathbf{C} \equiv \{dC_t\}_{t=0}^{\infty}$. In a stochastic model, such Jacobians can be interpreted as mappings between the coefficients of $MA(\infty)$ processes. (See Auclert et al. 2021.) For these processes to have finite variance,

[3] In this paper, we assume rational expectations, but Jacobians can also be modified to reflect departures from rational expectations (Auclert, Rognlie and Straub 2020). We show in appendix B.4 that many such modifications preserve quasi-Toeplitz structure.

it is necessary for their coefficients to be square-summable, in which case we can regard $\mathbf{M}$ as a mapping from the space of square-summable sequences ℓ^2 to itself. Formally, we say that $\mathbf{M}$ is a *bounded linear operator* on the real Hilbert space ℓ^2.[4] Note that this restriction to ℓ^2 is not without loss of generality: it excludes, for instance, the representative-agent model, which has a unit root in consumption, so that $\mathbf{M}$ maps to non-square-summable sequences.[5] For stationary models, however, ℓ^2 suffices.

In the following sections, we will consider a variety of sequence-space Jacobians $\mathbf{J}$, all understood as bounded linear operators on ℓ^2, and show that they have a very special "quasi-Toeplitz" structure. Any Jacobian $\mathbf{J} : \ell^2 \to \ell^2$ can be represented by a matrix $[J_{ts}]_{s=0,t=0}^{\infty}$, and we will refer to operators and their matrices interchangeably.

2.2 Toeplitz operators

We begin with an example. In the log-linearized Calvo model, the log reset price p_t^* chosen by a firm and the log aggregate price level p_t are determined in response to log nominal marginal cost mc_t according to the equations

$$p_t^* = (1-\beta\theta)\sum_{s=0}^{\infty}(\beta\theta)^s \mathbb{E}_t[mc_{t+s}] \tag{1}$$

$$p_t = (1-\theta)\sum_{s=0}^{\infty}\theta^s p_{t-s}^*. \tag{2}$$

In response to an MIT shock, for any variable x we have $\mathbb{E}_t[x_{t+s}] = x_{t+s}$ for $s \geq 0$ and $\mathbb{E}_t[x_{t+s}] = 0$ for $s < 0$. Given this, we can rewrite equations (1)–(2) as $\mathbf{p}^* = \mathbf{J}_{p^*,mc}\mathbf{mc}$ and $\mathbf{p} = \mathbf{J}_{p,p^*}\mathbf{p}^*$. Here, the Jacobian $\mathbf{J}_{p^*,mc}$ maps sequences of marginal costs $\mathbf{mc} = \{mc_t\}_{t=0}^{\infty}$ to sequences of reset prices $\mathbf{p}^* = \{p_t^*\}_{t=0}^{\infty}$, and the Jacobian $\mathbf{J}_{p,p^*}$ maps sequences of reset prices to sequences of aggregate prices $\mathbf{p} = \{p_t\}_{t=0}^{\infty}$. These Jacobians have upper (forward-

[4]Formally, $\mathbf{M}$ is the *Fréchet derivative* of $\mathcal{C}$ around the steady state, where $\mathcal{C} : \ell^2 \to \ell^2$ is recentered so that it is the mapping from deviations in income $\{Z_s - Z\}$ vs. the steady state to deviations in consumption $\{C_s - C\}$. All Fréchet derivatives are bounded linear operators. For useful introductory references to functional analysis, see Luenberger (1997) and Axler (2020).

[5]Formally, $\mathbf{M}$ maps to multiples of $\mathbf{1}$, which is not square-summable. This reflects the well-known fact that linearized representative-agent models can have a random walk component and thus infinite unconditional variance (Schmitt-Grohé and Uribe 2003).

looking) and lower (backward-looking) triangular forms, respectively:

$$\mathbf{J}_{p^*,mc} = (1-\beta\theta)\begin{pmatrix} 1 & \beta\theta & (\beta\theta)^2 & \\ 0 & 1 & \beta\theta & \ddots \\ 0 & 0 & 1 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix} \text{ and } \mathbf{J}_{p,p^*} = (1-\theta)\begin{pmatrix} 1 & 0 & 0 & \\ \theta & 1 & 0 & \ddots \\ \theta^2 & \theta & 1 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix}. \quad (3)$$

These Jacobians both belong to a special class of operators on ℓ^2, called *Toeplitz operators*.

Definition 1. *We say a linear operator* $\mathbf{J}$ *on* ℓ^2 *is* Toeplitz *if its matrix has constant entries along each diagonal*

$$\mathbf{J} = T(\mathbf{j}) \equiv \begin{pmatrix} j_0 & j_{-1} & j_{-2} & \\ j_1 & j_0 & j_{-1} & \ddots \\ j_2 & j_1 & j_0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix}, \quad (4)$$

where the $\mathbf{j} = \{j_k\}_{k=-\infty}^{\infty}$ *are absolutely summable,* $\sum_{k=-\infty}^{\infty} |j_k| < \infty$.

Lemma 1. *The linear operator* $T(\mathbf{j})$ *defined by* (4) *is bounded, with* $\|\mathbf{J}\| \leq \|\mathbf{j}\|_1$.[6]

Proof. See appendix B.1. ∎

If we plot the columns of a Toeplitz Jacobian and interpret them as impulse responses, they satisfy translation invariance with respect to time: the response at date t to a shock at date s is the same as the response at date $t+k$ to a shock at date $s+k$. Figures 2(a) and 2(b) depict this property for the Jacobians $\mathbf{J}_{p^*,mc}$ and $\mathbf{J}_{p,p^*}$ defined in (3).

By contrast, the mapping between mc_t and p_t—which combines equations (1) and (2)—involves the composition $\mathbf{J}_{p,p^*}\mathbf{J}_{p^*,mc}$. This *pass-through matrix* (Auclert, Rigato, Rognlie and Straub 2024c) is *not* Toeplitz, as evident from the lack of translation invariance from $s=0$ to $s=5$ in figure 2(c). We will return to this fact.

The Calvo model is hardly a special case. Since most simple macroeconomic equations are written in a time-invariant way, they will generally have Toeplitz representations in the sequence space. For instance, the log-linearized Euler equation $\mathbb{E}_t[r_{t+1}] = \sigma(\mathbb{E}_t[c_{t+1}] - c_t)$ can be written in terms of Toeplitz operators. To take an even simpler

[6]Formally, a linear operator $\mathbf{J}$ is bounded if there exists some $M \geq 0$ such that for all $\mathbf{x}$, $\|\mathbf{J}x\| \leq M\|x\|$. The minimum such M is the operator norm $\|\mathbf{J}\|$. Bounded operators are continuous in their inputs.

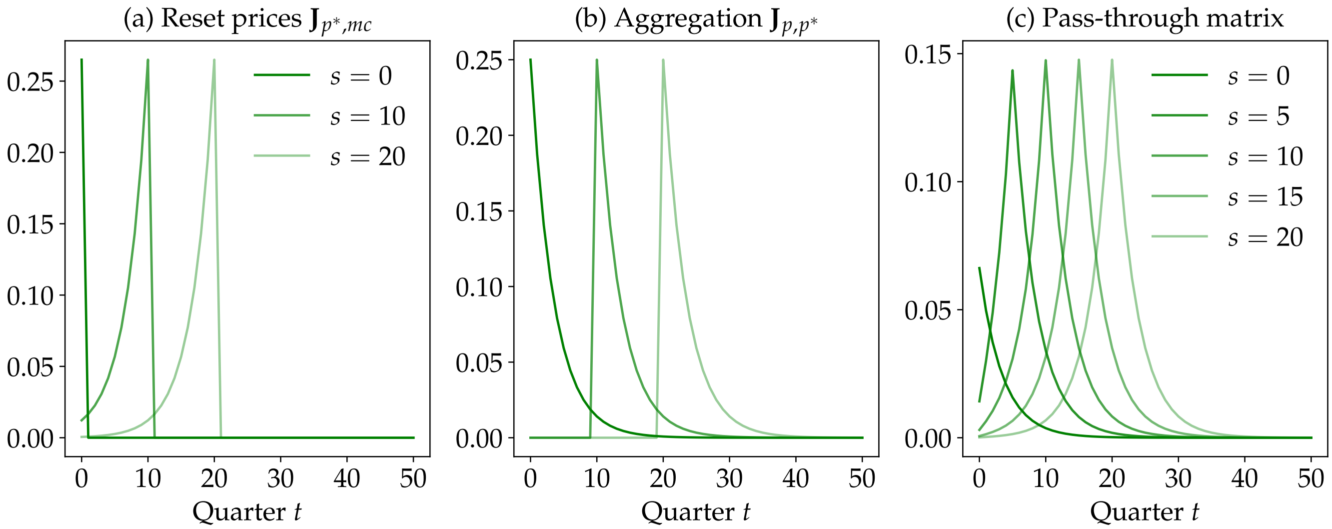


Figure 2: Columns of Calvo Model Jacobians

example, the familiar lead $\mathbf{F}$ and lag $\mathbf{L}$ operators are also Toeplitz:

$$\mathbf{F} \equiv \begin{pmatrix} 0 & 1 & 0 & \\ 0 & 0 & 1 & \ddots \\ 0 & 0 & 0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix} \qquad \mathbf{L} \equiv \begin{pmatrix} 0 & 0 & 0 & \\ 1 & 0 & 0 & \ddots \\ 0 & 1 & 0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix}. \tag{5}$$

In definition 1, we assume that $\mathbf{j}$ is a scalar sequence, but more generally it can be a matrix sequence, in which case $T(\mathbf{j})$ is called *block Toeplitz*. We develop this case formally in section 3.7. The intuition and results we derive with Toeplitz operators will carry over to the block case with few modifications.

Symbols. Each Toeplitz operator $T(\mathbf{j})$ is defined by a two-sided sequence $\mathbf{j}$. This may be transformed into another object, the *symbol*, which will prove to be particularly useful when we multiply and invert Toeplitz operators.

Definition 2. *The* symbol *of a Toeplitz operator $T(\mathbf{j})$ is the complex-valued function j over the unit circle $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$, $j : \mathbb{T} \to \mathbb{C}$, given by the Laurent series*

$$j(z) \equiv \sum_{k=-\infty}^{\infty} j_k z^k. \tag{6}$$

This series always converges, since we assumed that the j_k are absolutely summable.

For Toeplitz $\mathbf{J}$, evaluating the matrix-vector product $\mathbf{y} = \mathbf{J}\mathbf{x}$ is equivalent to multiplying the symbol $j(z)$ and $x(z) \equiv \sum_{k=0}^{\infty} x_k z^k$ to obtain a new complex-valued function $y(z)$ on $\mathbb{T}$, and then defining the output sequence $\mathbf{y}$ to be the coefficients on nonnegative powers in the Laurent expansion of $y(z)$. This connects our analysis to the early literature on rational expectations models in the frequency domain, including Hansen and Sargent (1980b,a) and Whiteman (1985), where such operations were common.[7]

2.3 Quasi-Toeplitz Jacobians: missing anticipation, compact correction

Is the product of Toeplitz operators still Toeplitz? Generally not. Take the lead and lag operators $\mathbf{F}$ and $\mathbf{L}$. One can verify that $\mathbf{FL}$ is the identity $\mathbf{I}$, which is Toeplitz. But $\mathbf{LF}$ is not the identity. Instead, it is

$$\mathbf{LF} = \begin{pmatrix} 0 & 0 & 0 & \\ 0 & 1 & 0 & \ddots \\ 0 & 0 & 1 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix}, \tag{7}$$

which is almost the identity, but with the $(0, 0)$ entry replaced by a 0.[8] As a result, this matrix is not Toeplitz.

Where did the non-Toeplitz part come from? If there is an unanticipated shock at date 0, then the lagged expectation at date 0—which is the expectation at date -1 of x_0, $\mathbb{E}_{-1}[x_0]$—is zero. We call this *missing anticipation*, and it happens whenever we take the lag of a lead.

Now consider the case of general Toeplitz operators $T(\mathbf{a})$ and $T(\mathbf{b})$. Their product $\mathbf{C} \equiv T(\mathbf{a})T(\mathbf{b})$ has entries

$$C_{s+j,s} = \sum_{u=0}^{\infty} a_{s+j-u} b_{u-s} \tag{8}$$

Substituting $v \equiv u - s$ in the sum and taking the limit as $s \to \infty$:

$$C_{s+j,s} = \sum_{v=-s}^{\infty} a_{j-v} b_v \longrightarrow \sum_{v=-\infty}^{\infty} a_{j-v} b_v \equiv c_j \tag{9}$$

[7]Formally, we can regard a Toeplitz operator $T(\mathbf{j})$ as an operator on the "Hardy space" $H^2(\mathbb{T})$, which is the space of analytic functions—whose Laurent series have no negative powers—on the unit circle. For any analytic function $x(z) = \sum_{k=0}^{\infty} x_k z^k$, the Toeplitz operator $T(\mathbf{j})$ multiplies by $j(z)$ and then *projects* back onto the Hardy space by dropping all the negative powers in the Laurent series. This characterization of Toeplitz operators—as the composition of projection and multiplication—is a common viewpoint in the mathematical literature, but for our purposes it is useful to work more explicitly with matrices and sequences.

[8]$\mathbf{F}$ maps $(x_0, x_1, x_2, \ldots)$ to $(x_1, x_2, \ldots)$, and then $\mathbf{L}$ maps this to $(0, x_1, x_2, \ldots)$, replacing the x_0 with a 0.

It follows that asymptotically, $\mathbf{C}$ looks Toeplitz, with entries $\mathbf{c} = \{c_j\}$ given by the *convolution* of $\mathbf{a}$ and $\mathbf{b}$. The corresponding symbol $c(z)$ is simply the product $a(z)b(z)$, which follows from the convolution theorem for z-transforms.

To illustrate this, consider the Calvo pricing example from (3), and take the product of $T(\mathbf{a}) \equiv \mathbf{J}_{p,p^*}$, which maps reset prices to aggregate prices, with $T(\mathbf{b}) \equiv \mathbf{J}_{p^*,mc}$, which maps marginal costs to reset prices. As we can see in figure 2(c), the resulting "pass-through" matrix $\mathbf{C} = T(\mathbf{a})T(\mathbf{b})$ from marginal cost to aggregate prices is not Toeplitz: in early periods, the reaction of prices to marginal cost shocks is less, because pricesetters before date 0 do not anticipate the shocks. Eventually, however, the columns of the pass-through matrix converge to Toeplitz form: visually, the responses to a $s = 10$ and $s = 20$ shock are almost identical, just shifted.

The effects of missing anticipation. Writing $t \equiv s + j$, (9) implies the following recursion for all $t, s > 0$, relating $C_{t,s}$ to the previous entry $C_{t-1,s-1}$ on the same diagonal:

$$C_{t,s} = C_{t-1,s-1} + a_t b_{-s} \tag{10}$$

In our pricing example, $b_{-s} = (1 - \beta\theta)(\beta\theta)^s$ and $a_t = (1 - \theta)\theta^t$. Consider, for instance, the response $C_{4,3}$ of aggregate prices at date 4 to a marginal cost shock at date 3. Equation (10) says that this is $C_{3,2}$ plus the additional term $a_4 b_{-3}$. This term reflects the effect of the marginal cost shock on date-0 reset prices (b_{-3}), times the effect of these on date-4 aggregate prices (a_4). This term is not in $C_{3,2}$, because no one anticipates a date-2 shock 3 periods in advance.

As we move down the diagonals of $\mathbf{C}$, we add more terms of the form $a_t b_{-s}$ until convergence. The difference $\mathbf{E} \equiv \mathbf{C} - T(\mathbf{c})$ between $\mathbf{C}$ and its Toeplitz limit is given by the remaining terms on the same diagonal:

$$E_{t,s} \equiv C_{t,s} - c_{t-s} = -\sum_{k=1}^{\infty} a_{t+k} b_{-s-k} \tag{11}$$

Economically, we can think of c_j as giving the effect, at an offset of j, of an infinitely well-anticipated shock. But if the shock happens at some finite date s, then (11) tells us that we need to subtract off any anticipatory effects with a horizon of $s + 1$ or more, because the shock was unanticipated prior to date 0. As t and s become higher, the missing anticipation in (11) becomes smaller and smaller, and $E_{t,s}$ goes to zero. For instance, in our pricing example, the terms $a_t b_{-s}$ are proportional to $\beta^s \theta^{s+t}$, decaying exponentially.

Indeed, it is possible to show a stronger result: that the operator $\mathbf{E}$ is *compact* in ℓ^2.[9] We summarize our results thus far in the following proposition.

Proposition 1. *The product of any two Toeplitz operators $T(\mathbf{a})$ and $T(\mathbf{b})$ is*

$$T(\mathbf{a})T(\mathbf{b}) = T(\mathbf{c}) + \mathbf{E} \tag{12}$$

where $\mathbf{c} \equiv \mathbf{a} * \mathbf{b}$ is the convolution of $\mathbf{a}$ and $\mathbf{b}$, with $c(z) = a(z)b(z)$, and $\mathbf{E}$ is the compact operator defined by (11).

Proof. See appendix B.1. ■

This proposition is encouraging. Even though the product of two Toeplitz operators is not usually Toeplitz, it belongs to a more general class of highly structured operators: *quasi-Toeplitz operators*, defined to take the form on the right in (12).[10]

Definition 3. *We say a bounded linear operator $\mathbf{J}$ on ℓ^2 is* quasi-Toeplitz *if it can be written as*

$$\mathbf{J} = T(\mathbf{j}) + \mathbf{E} \tag{13}$$

where $T(\mathbf{j})$ is Toeplitz and $\mathbf{E}$ is a compact operator on ℓ^2 known as the (compact) correction.

Indeed, this more general class is closed under multiplication.

Proposition 2. *The product of any two quasi-Toeplitz operators is quasi-Toeplitz.*

Proof. For any $\mathbf{A}_1 = T(\mathbf{a}_1) + \mathbf{E}_1$ and $\mathbf{A}_2 = T(\mathbf{a}_2) + \mathbf{E}_2$, we compute $\mathbf{A}_1\mathbf{A}_2 = T(\mathbf{a}_1)T(\mathbf{a}_2) + \mathbf{E}_1 T(\mathbf{a}_2) + T(\mathbf{a}_1)\mathbf{E}_2 + \mathbf{E}_1\mathbf{E}_2$. The first term is quasi-Toeplitz by proposition 1, and the other terms are products of bounded and compact operators and thus compact. ■

This property suggests that we will often see quasi-Toeplitz operators when solving models to first order in the sequence space, because this often involves composing Jacobians that are Toeplitz—because the underlying equations are time-invariant, like in (3)—with each other.

2.4 Connection to fake news matrix

The *fake news matrix* was introduced by Auclert et al. (2021) as part of a fast algorithm to obtain sequence-space Jacobians for heterogeneous-agent models. For an arbitrary Jacobian $\mathbf{J}$, we define this matrix as follows.

[9]There are several equivalent ways to define compactness on a Hilbert space. One is that a compact operator can be arbitrarily well-approximated, in the operator norm, by finite-rank operators.

[10]There is a growing literature on quasi-Toeplitz operators. See, e.g., Bini et al. (2019).

Definition 4. *The* fake news matrix $\boldsymbol{\mathcal{F}}$ *corresponding to a Jacobian* $\mathbf{J}$ *is defined as*

$$\mathcal{F}_{t,s} = \begin{cases} J_{t,s} - J_{t-1,s-1} & t,s > 0 \\ J_{t,s} & t = 0 \text{ or } s = 0. \end{cases} \tag{14}$$

The first column of $\boldsymbol{\mathcal{F}}$ is the same as $\mathbf{J}$, but later columns can be interpreted as giving the pure effect of anticipation: concretely, for $s > 0$, $\mathcal{F}_{t,s}$ gives the effect at date t from having anticipated at date 0 that a shock will happen at date s.[11]

As the following proposition shows, there is a close connection between quasi-Toeplitz structure and the fake news matrix. If we have a quasi-Toeplitz $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$, we can express its components $\mathbf{j}$ and $\mathbf{E}$ easily in in terms of $\boldsymbol{\mathcal{F}}$. Given any absolutely summable $\boldsymbol{\mathcal{F}}$, we can also build a quasi-Toeplitz $\mathbf{J}$.

Proposition 3. *Suppose that* $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ *is quasi-Toeplitz, and let* $\boldsymbol{\mathcal{F}}$ *be the corresponding fake news matrix. Then the kth entry* j_k *of the symbol is the sum of entries of* $\boldsymbol{\mathcal{F}}$ *on the kth lower diagonal, and each entry* $E_{t,s}$ *of the correction is minus the sum of entries after* (t,s) *on its diagonal:*

$$j_k \equiv \sum_{v=\max\{0,-k\}}^{\infty} \mathcal{F}_{k+v,v} \qquad \text{and} \qquad E_{t,s} \equiv -\sum_{k=1}^{\infty} \mathcal{F}_{t+k,s+k}. \tag{15}$$

Conversely, for any $\boldsymbol{\mathcal{F}}$ *whose entries have a finite absolute sum, the corresponding Jacobian* $\mathbf{J}$ *given by the recursion* $J_{t,s} = J_{t-1,s-1} + \mathcal{F}_{t,s}$ *for* $t,s > 0$ *and* $J_{t,s} = \mathcal{F}_{t,s}$ *for* $t = 0$ *or* $s = 0$ *is quasi-Toeplitz, with symbol and correction given by* (15).

Proof. See appendix B.2. ∎

Intuitively, j_k is the response, at a lag of k periods, to an infinitely well-anticipated shock, and we sum the effect of this anticipation at all previous periods by summing entries of the fake news matrix. Meanwhile, $E_{t,s}$ is the effect of missing anticipation—the fact that the shock at s was not anticipated more than s periods in advance—and we obtain it by negating the sum of fake news entries that have horizons longer than s.

In our Calvo pricing example, the fake news matrix corresponding to the pass-through matrix $\mathbf{C}$ is $\mathcal{F}_{t,s} = (1-\theta)\theta^t \cdot (1-\beta\theta)(\beta\theta)^s$, as we can read off from (10). The second factor $(1-\beta\theta)(\beta\theta)^s$ is the anticipatory response of reset prices to a marginal cost increase in s periods, while the first factor $(1-\theta)\theta^t$ gives the persistent effect of these reset prices on aggregate prices. This is a special case of a general expression, derived in appendix B.3, for

[11]For $s > 0$, $\mathcal{F}_{t,s}$ can also be interpreted as the response at date t to a “fake news” shock: news about a change in input at date s that is announced at date 0 and retracted at date 1. To first order, the effect of such a shock is equal to the effect of anticipation at date 0.

the fake news matrix corresponding to any product of two Toeplitz matrices. Our earlier expression (11) for the correction $\mathbf{E}$ in this product is a special case of (15).

In the next two subsections, we will apply proposition 3 to prove *structure theorems*, showing that broad classes of sequence-space Jacobians are quasi-Toeplitz. Although for simplicity we assume rational expectations here, appendix B.4 shows that the quasi-Toeplitz result in proposition 3 also extends to many departures from full information and rational expectations.

2.5 Structure theorem for heterogeneous-agent blocks

Sequence-space Jacobians are particularly valuable for heterogeneous-agent models, because they allow us to bypass potentially high-dimensional heterogeneity and solve for equilibrium entirely in terms of aggregate sequences. Auclert et al. (2021) provide a fast algorithm to obtain Jacobians for *heterogeneous-agent blocks*, which we define following that paper.

Definition 5. *A* heterogeneous-agent block *consists of three equations*

$$\mathbf{v}_t = v(\mathbf{v}_{t+1}, X_t) \tag{16}$$

$$\mathbf{D}_{t+1} = \Lambda(\mathbf{v}_{t+1}, X_t)'\mathbf{D}_t \tag{17}$$

$$Y_t = y(\mathbf{v}_{t+1}, X_t)'\mathbf{D}_t, \tag{18}$$

where $\mathbf{v}$, $\mathbf{D}$, and $\mathbf{y}$ are finite-dimensional vectors, and X_t and Y_t are scalars. A steady state *of this block solves (16)–(18) for constant X. Dynamically, the block maps an input sequence $\mathbf{X} = \{X_k\}_{k=0}^{\infty}$ to an output sequence $\mathbf{Y} = \{Y_k\}_{k=0}^{\infty}$, given some initial $\mathbf{D}_0$. Assuming local differentiability, we denote the sequence-space Jacobian of $\mathbf{Y}$ with respect to $\mathbf{X}$ around the steady state by $\mathbf{J}$.*

This system of equations encapsulates three conceptually distinct steps in evaluating a heterogeneous-agent block:

- Equation (16) is a *backward iteration*: it captures how $\mathbf{v}_t$—some representation of the value function—is determined by future $\mathbf{v}_{t+1}$ as well as the current input X_t;
- Equation (17) is a *forward iteration*: it captures how the transition matrix Λ between today's distribution $\mathbf{D}_t$ and tomorrow's distribution $\mathbf{D}_{t+1}$ is determined by next period's $\mathbf{v}_{t+1}$ as well as today's input X_t;
- Equation (18) is a *measurement equation*: it captures how $\mathbf{v}_{t+1}$ and $\mathbf{D}_t$, together with today's input X_t, determine today's output Y_t of interest.

For instance, consider the household block of a standard incomplete markets model, like in Aiyagari (1994) or Krusell and Smith (1998). Holding wages fixed, households take as an input a sequence of asset returns $\mathbf{x} = \{r_t\}_{t=0}^{\infty}$ and make consumption-savings decisions, which result in a sequence $\mathbf{y}_t = \{A_t\}_{t=0}^{\infty}$ of aggregate asset holdings.

There is also a subclass of heterogeneous-agent blocks of particular interest, which we call *stationary* heterogeneous-agent blocks.

Definition 6. *A heterogeneous-agent block is* stationary *if:*

1. *the Jacobian of $\mathbf{v}_t$ with respect to $\mathbf{v}_{t+1}$ around the steady state has all eigenvalues inside the unit circle;*

2. *the steady-state transition matrix $\mathbf{\Lambda}_{ss}$ has all eigenvalues except one inside the unit circle.*

These conditions are quite loose, and hold for almost all heterogeneous-agent blocks of interest. Condition 1 is usually guaranteed by discounting, and is necessary for the backward iteration to converge to steady-state value and policy functions. Condition 2 says that there is a unique stationary distribution across idiosyncratic states.[12]

Condition 1 implies that the effect of a shock on today's value and policy functions decays as the shock moves further into the future. Condition 2 implies that the effect of a change in today's distribution on the future distribution decays as we move further into the future. Together, these observations imply a bound on entries of the fake-news matrix $\boldsymbol{\mathcal{F}}$ corresponding to $\mathbf{J}$.

Lemma 2. *For a stationary heterogeneous-agent block, we have $|\mathcal{F}_{t,s}| \leq C\Delta^{s+t}$ for some $\Delta \in (0,1)$ and $C > 0$.*

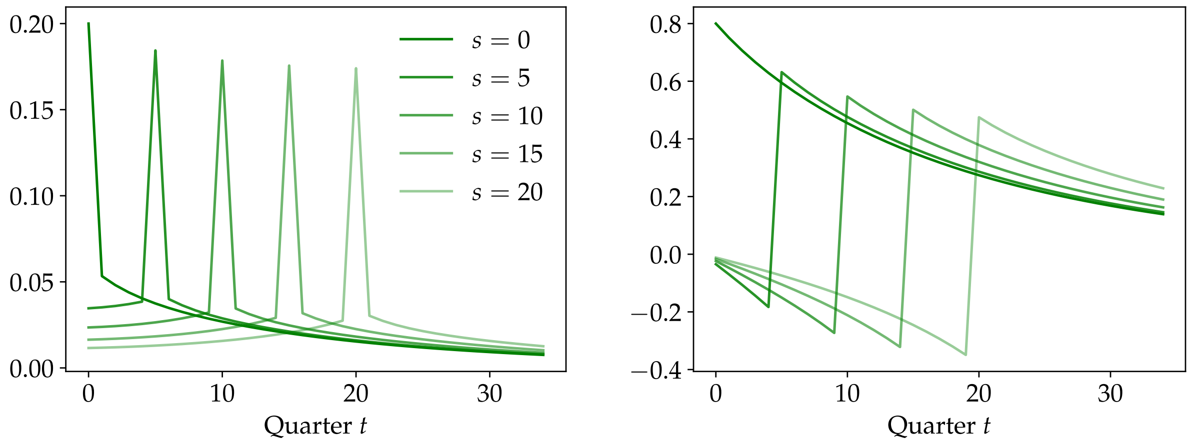
Proof. See appendix B.5. ■

Here, Δ is some scalar less than 1 and greater than all non-unit eigenvalues discussed in definition 6. This bound on entries of $\boldsymbol{\mathcal{F}}$ implies that they are absolutely summable, and it follows immediately from proposition 3 that $\mathbf{J}$ is quasi-Toeplitz.

Proposition 4. *The Jacobian $\mathbf{J}$ of a stationary heterogeneous-agent block is quasi-Toeplitz.*

Since the algorithm in Auclert et al. (2021) for computing $\mathbf{J}$ in heterogeneous-agent blocks builds the fake news matrix $\boldsymbol{\mathcal{F}}$ as an intermediate object, it is easy to obtain the quasi-Toeplitz symbol and correction from (15). See appendix A.2 for additional details on the symbol.

[12]If there is permanent heterogeneity, then condition 2 does not hold, but this situation can easily be handled by separating different permanent types into different heterogeneous-agent blocks, which are individually stationary.



(a) Consumption Jacobian **M** (iMPCs)

(b) Asset Jacobian **A**

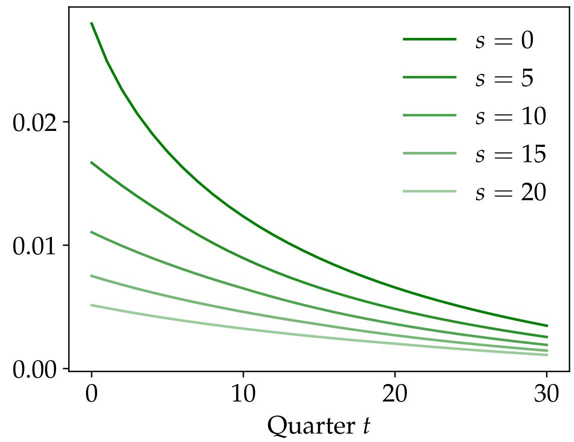
Figure 3: Columns of HA Jacobians with respect to aggregate income

Example: standard incomplete markets household block. Here, we suppose that the household block is given by a standard incomplete markets model, where households consume and save subject to idiosyncratic income risk and a borrowing constraint. The calibration is described in appendix B.6.

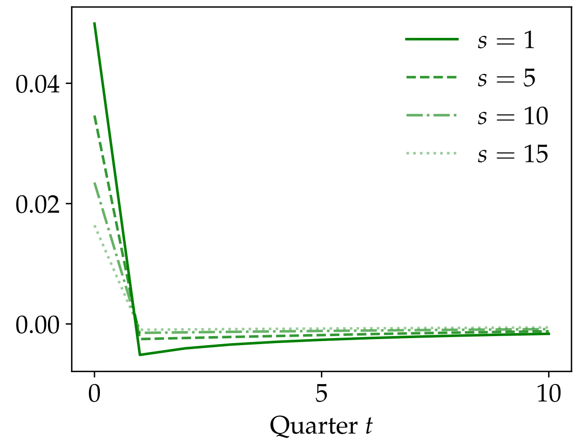
In figure 3, we plot columns of two Jacobians for this model, both of which are discussed at length in Auclert et al. (2024a). First, we have **M**, which we briefly discussed in section 2.1. This is the Jacobian of aggregate consumption with respect to a shock that shifts all households' incomes proportionally. We call the entries of **M** intertemporal marginal propensities to consume (iMPCs). Second, we have the Jacobian **A**, giving the response of aggregate assets to the same shock.

The quasi-Toeplitz structure of **M** and **A** is clear, as both have columns that converge to a limiting two-sided sequence. For **M**, we note that there is a stronger consumption response out of an unanticipated shock at $s = 0$ vs. anticipatory shocks at later s, because households come in with more assets at s when they have not done anticipatory spending. As we see in **A**, this anticipatory spending results in draw-down of assets prior to s.

In figure 4(a), we plot columns of the compact correction $\mathbf{E} \equiv \mathbf{M} - T(\mathbf{m})$ of **M**. We see that **E** is positive: missing anticipation means that spending is higher, because assets are drawn down less ahead of time. In figure 4(b), we plot columns of the fake news matrix $\boldsymbol{\mathcal{F}}$ corresponding to **M** (omitting column $s = 0$, which is the same as column 0 of **M** in figure 3(a)). For all $s > 0$, $\mathcal{F}_{t,s}$ is positive at $t = 0$ and negative afterward: the effect of anticipating future income at s is to spend more now, leaving less for the future. Importantly, entries of $\boldsymbol{\mathcal{F}}$ decay rapidly, illustrating lemma 2.



(a) Compact correction $\mathbf{M} - T(\mathbf{m})$



(b) Fake news matrix $\boldsymbol{\mathcal{F}}$

Figure 4: Columns of correction and fake news matrices for $\mathbf{M}$

Following Auclert et al. (2024a), we can use these Jacobians to characterize the output response to a fiscal shock, assuming that monetary policy holds real interest rates constant. First, the *intertemporal Keynesian cross* is a statement of goods market clearing: $d\mathbf{Y} = d\mathbf{G} - \mathbf{M}d\mathbf{T} + \mathbf{M}d\mathbf{Y}$, where $d\mathbf{G}$ and $d\mathbf{T}$ are shocks to spending and taxes, and $d\mathbf{Y}$ is the resulting change in equilibrium output. There is also an equivalent asset market clearing formulation, which can be written as $\mathbf{A}d\mathbf{Z} = d\mathbf{B}$, where $d\mathbf{Z} \equiv d\mathbf{Y} - d\mathbf{T}$ is after-tax income, and $d\mathbf{B}$ is the change in bond supply from the fiscal shock. We will discuss how to solve the latter, and verify determinacy and existence of solutions, in section 3.

2.6 Structure theorem for expectational linear difference equations

Dynamic macroeconomic models, once linearized, are typically written as expectational linear difference equations. In this section, we show that the mappings between sequences implied by such equations are quasi-Toeplitz.

Simple example. For a simple illustration, consider the following scalar equation:

$$a_{-1}\mathbb{E}_t[x_{t+1}] + a_0 x_t + a_1 x_{t-1} + b_0 u_t = 0. \tag{19}$$

Here, x_t is an endogenous variable determined at date t, and u_t is an exogenous variable following some stochastic process.

If $a_{-1}\lambda^2 + a_0\lambda + a_1$ has one stable root $|p| < 1$ and one unstable root $|q| > 1$, by standard arguments there exists a unique solution for x_t in terms of the endogenous state

x_{t-1} and expected exogenous u_{t+j}, which can be written

$$x_t = px_{t-1} + \frac{b_0}{a_{-1}q} \sum_{j=0}^{\infty} q^{-j} \mathbb{E}_t[u_{t+j}]. \tag{20}$$

Equation (20) is a state-space law of motion for x_t. To find its sequence-space counterpart, which maps impulses to u to impulses to x, we consider an MIT shock, where $u_s = 0$ for $s < 0$, and then the impulse $\mathbf{u} \equiv \{u_0, u_1, \ldots\}$ becomes known at date 0. The mapping from $\mathbf{u}$ to $\mathbf{x} \equiv \{x_0, x_1, \ldots\}$ implied by (20) is then

$$\mathbf{x} = \underbrace{T(\mathbf{p})T(\mathbf{q})}_{\equiv \mathbf{J}_{x,u}} \mathbf{u},$$

where $T(\mathbf{p})$ and $T(\mathbf{q})$ are Toeplitz matrices with symbols $p(z) = \sum_{j=0}^{\infty} p^j z^j$ and $q(z) = \frac{b_0}{a_{-1}q} \sum_{j=0}^{\infty} q^{-j} z^{-j}$, respectively. Here, $T(\mathbf{q})$ maps $\mathbf{u}$ to the sequence given by the right term in (20), and then $T(\mathbf{p})$ accumulates that sequence to obtain $\mathbf{x}$. The composite Jacobian $\mathbf{J}_{x,u}$, as the product of two Toeplitz matrices, is quasi-Toeplitz. Indeed, since $\mathbf{p}$ decays exponentially, it is possible to show that the compact correction for $\mathbf{J}_{x,u}$ has rank one.

Why is $\mathbf{J}_{x,u}$ so close to being exactly Toeplitz, separated only by a rank-one matrix? The intuition is that missing anticipation filters through the state variable x_{t-1}. As a result, the correction, which captures this missing anticipation, has to have all columns be proportional to the impulse response of $\{x_{t+h}\}$ to x_{t-1}, i.e. proportional to $\{p, p^2, p^3, ...\}$.

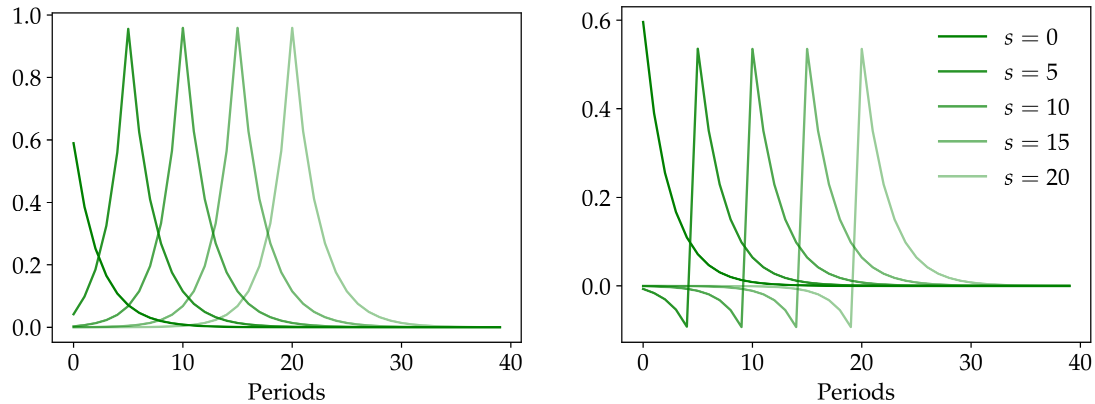
Structure theorem for the general case. We now specify a more general, multivariate expectational linear difference equation in canonical form

$$\mathbf{A}\mathbb{E}_t[x_{t+1}] + \mathbf{B}x_t + \mathbf{C}x_{t-1} + \mathbf{D}u_t = \mathbf{0}_{n\times 1}. \tag{21}$$

Here, x_t is an $n \times 1$ vector of endogenous variables and u_t is an $m \times 1$ vector of exogenous variables.[13] The vast majority of models of interest to us, once linearized, may be written in such a way. The following result generalizes the findings from our simple example above.

Proposition 5. *Suppose that the system in (21) satisfies the conditions (stated in appendix B.8) for a unique, stable state-space solution. Then, the sequence-space Jacobians that map each exogenous*

[13] This is the same as the form in Uhlig (1995), except that for simplicity we omit expectations of u_{t+1} (which, since they are exogenous, can be incorporated into u_t itself), and we do not require an explicit recursive law of motion for u_t.



(a) Capital to u

(b) Capital to TFP

Figure 5: Columns of Jacobians of capital to u and TFP

variable to each endogenous variable will be quasi-Toeplitz and the ranks of the correction matrices will be, at most, the size of the state space.

Proof. See appendix B.7. ∎

Proposition 5 implies that quasi-Toeplitz matrices are ubiquitous. Under standard conditions for the linearized solution to a dynamic macro model, the Jacobians characterizing the sequence-space solution are *all* quasi-Toeplitz.[14] The only limitation is stationarity—the solution cannot have unit roots.

Example: neoclassical growth model. To apply this result, we consider a standard neoclassical growth model, where a representative agent allocates production $f(K_{t-1}, Z_t)$ to capital and consumption $K_t + C_t$ in each period to maximize intertemporal utility, given by the expected discounted sum of $U(C_t)$. (See appendix B.9 for details.) Linearized around the steady state, the difference equation defining this model is

$$f_K dK_{t-1} - \left(1 + f_K + \frac{f_{KK}}{f_K}\frac{U'}{U''}\right) dK_t + \mathbb{E}_t[dK_{t+1}] = \left(f_Z + \frac{f_{KZ}}{f_K}\frac{U'}{U''}\right) \mathbb{E}_t[dZ_{t+1}] - f_Z dZ_t. \tag{22}$$

Letting $x_t \equiv dK_t$ and u_t be minus the right side of (22), this system is of the form (19), and we conclude that the Jacobian $\mathbf{J}_{dK,u}$ mapping u_t to dK_t is quasi-Toeplitz with a correction of rank 1. This Jacobian is visualized in panel (a) of figure 5.

[14]Indeed, the structure theorem for heterogeneous-agent blocks in the previous section can be interpreted as one special case of this very general result, when (16)–(18) are linearized and cast into the form (21).

To obtain the Jacobian of capital to TFP, we multiply $\mathbf{J}_{dK,u}$ by the Jacobian mapping dZ to u, which is Toeplitz. The resulting Jacobian $\mathbf{J}_{dK,dZ}$, as the product of quasi-Toeplitz and Toeplitz matrices, remains quasi-Toeplitz.[15] This Jacobian is visualized in figure 5(b), where we see that it resembles the asset Jacobian in figure 3(b), with depletion of capital in anticipation of a TFP shock, and then a large increase in capital in the period of the shock, which is drawn down over time.

Solving directly in the sequence space. In this section, we have built upon standard solution methods that deliver a state-space law of motion, deriving sequence-space Jacobians from these. An alternative approach is to solve entirely in the sequence space.

For instance, in our neoclassical growth model example, we can stack (19) for all $t = 0, 1, \ldots$, defining $\mathbf{H_K}$ to be the Toeplitz matrix mapping $d\mathbf{K} \equiv \{dK_0, dK_1, \ldots\}$ to the stacked left side of (19), and $-\mathbf{H_Z}$ to be the Toeplitz matrix mapping $d\mathbf{Z}$ to the right. The sequence-space system is then $\mathbf{H_K}d\mathbf{K} = -\mathbf{H_Z}d\mathbf{Z}$, and the Jacobian $\mathbf{J}_{dK,dZ}$ in figure 5(b) equals $-\mathbf{H_K}^{-1}\mathbf{H_Z}$.[16] The next section discusses when such an inverse $\mathbf{H_K}^{-1}$, which guarantees a unique solution, actually exists.

3 Determinacy, existence, and invertibility

So far, we have dealt with sequence-space Jacobians $\mathbf{J}$ as mappings between sequences. To solve for equilibrium, however, it is often necessary to solve a sequence-space *system* $\mathbf{Jx} = \mathbf{y}$ for $\mathbf{x}$ given $\mathbf{y}$. Examples from the previous section include the asset version of the intertemporal Keynesian cross, $\mathbf{A}d\mathbf{Z} = d\mathbf{B}$, and the system for the neoclassical growth model, $\mathbf{H_K}d\mathbf{K} = -\mathbf{H_Z}d\mathbf{Z}$.

If $\mathbf{J}$ is invertible, then $\mathbf{Jx} = \mathbf{y}$ has a unique solution $\mathbf{x} = \mathbf{J}^{-1}\mathbf{y}$. Alternatively, there may be some $\mathbf{y}$ for which there is no solution $\mathbf{x}$. We denote the dimension of the subspace of such $\mathbf{y}$ by $\text{nonex}(\mathbf{J}) \equiv \dim((\text{range } \mathbf{J})^{\perp})$. There may also be multiple solutions $\mathbf{x}$ for a given $\mathbf{y}$. We denote the dimension of the affine subspace of such solutions by $\text{indet}(\mathbf{J}) \equiv$

[15] Indeed, as appendix B.9 shows, its correction remains rank one, since there is still only one state variable, capital dK_{t-1}, and any missing anticipation of dZ feeds in through dK_{t-1}. The same is true for the quasi-Toeplitz matrix mapping dZ to output dY.

[16] An alternative approach would be to construct a larger sequence-space system, with some Jacobians representing the supply side of the model and others representing the map from incomes and interest rates to household consumption or assets, with a market-clearing condition linking the two. This approach works well when there is a heterogeneous-agent household side (see, e.g., the "Krusell-Smith" model in Auclert et al. 2021). But in the representative-agent case, as mentioned earlier in this section, household Jacobians do not map ℓ^2 to itself, and our machinery does not directly apply, even though the general equilibrium Jacobians are ultimately stationary.

dim(null $\mathbf{J}$).[17] In this section, we show how to characterize these for Toeplitz and quasi-Toeplitz $\mathbf{J}$, building upon the winding number test introduced by Onatski (2006).

3.1 The winding number criterion for Toeplitz operators

We defined the symbol $j(z)$ of a Toeplitz operator $T(\mathbf{j})$ earlier as a complex-valued function on the unit circle. As for any such complex function, as long as it has no zeros on the unit circle, we can calculate its *winding number*.

Definition 7. *The winding number* $\mathrm{wind}(j)$ *of a continuous function* $j : \mathbb{T} \to \mathbb{C} \setminus \{0\}$ *is*

$$\mathrm{wind}(j) = \frac{1}{2\pi i} \oint_j \frac{dz}{z}, \tag{23}$$

which is equal to the number of times the graph of $j(z)$ goes counterclockwise around the origin as z goes counterclockwise along $\mathbb{T}$.

It turns out that the winding number offers a simple way to determine whether a Toeplitz operator $\mathbf{J} = T(\mathbf{j})$ is invertible—and if not, in which way it fails to be invertible.

Proposition 6. *Consider a Toeplitz operator $\mathbf{J} = T(\mathbf{j})$ with symbol $j(z) = \sum_{k=-\infty}^{\infty} j_k z^k$ such that $j(z) \neq 0$ for $z \in \mathbb{T}$. Then:*

1. *If* $\mathrm{wind}(j) = 0$, $\mathbf{J}$ *is invertible, and its inverse* $\mathbf{J}^{-1}$ *is quasi-Toeplitz.*
2. *If* $\mathrm{wind}(j) < 0$, $\mathbf{Jx} = \mathbf{y}$ *has indeterminacy of dimension* $\mathrm{indet}(\mathbf{J}) = -\mathrm{wind}(j)$, *but a solution always exists,* $\mathrm{nonex}(\mathbf{J}) = 0$.
3. *If* $\mathrm{wind}(j) > 0$, $\mathbf{Jx} = \mathbf{y}$ *has nonexistence of dimension* $\mathrm{nonex}(\mathbf{J}) = \mathrm{wind}(j)$, *but all solutions are unique,* $\mathrm{indet}(\mathbf{J}) = 0$.

Proof. See appendix C.1. ■

Remarkably, for Toeplitz operators, we never have *both* indeterminacy and nonexistence. This is the opposite of the situation in finite-dimensional linear algebra, where indeterminacy and nonexistence necessarily go together.

Why is $\mathbf{J}^{-1}$, if it exists, quasi-Toeplitz? Suppose we take $T(\mathbf{j}^{-1})$, where $\mathbf{j}^{-1}$ is the inverse of $\mathbf{j}$ satisfying $\mathbf{j}^{-1} * \mathbf{j} = 1$.[18] This is *almost* an inverse, because the Toeplitz part of the

[17] We say a bounded operator $\mathbf{J}$ is *invertible* if there is a bounded linear operator $\mathbf{J}^{-1}$ satisfying $\mathbf{J}\mathbf{J}^{-1} = \mathbf{J}^{-1}\mathbf{J} = \mathbf{I}$. In Hilbert space, this is true whenever $\mathbf{J}$ is bijective, so if $\mathrm{nonex}(\mathbf{J}) = 0$ and $\mathrm{indet}(\mathbf{J}) = 0$, then $\mathbf{J}$ is invertible.

[18] The '1' here refers to the two-sided sequence with 1 at position zero and 0s elsewhere.

product $T(\mathbf{j}^{-1})T(\mathbf{j})$ is the identity. But this product will also generally have a compact correction, and to offset this, the inverse $\mathbf{J}^{-1} = T(\mathbf{j}^{-1}) + \mathbf{E}$ must include some compact correction $\mathbf{E}$ as well.

Numerical implementation. See appendix A.1 for details on how to numerically implement the winding number criterion. A few key points stand out. First, if a Jacobian is defined as the product or sum of other Jacobians, then it is best to evaluate the symbol for those Jacobians instead: e.g. if $\mathbf{J} \equiv \mathbf{J}_1\mathbf{J}_2$, then we can calculate $j_1(z)$ and $j_2(z)$ individually, then multiply to obtain $j(z)$. Second, the fast Fourier transform makes calculating $j(z)$ much more efficient. Finally, the entire Jacobian is not necessary to apply the criterion: all we need is the sequence $\ldots, j_{-1}, j_0, j_1, \ldots$, which can be obtained numerically as the response to a very well-anticipated shock.

Simple applications of the winding number criterion. For the lag operator $\mathbf{L}$, the symbol is z and its graph as a function of the angle θ is $\cos\theta + \sin\theta \cdot i$, which has a winding number of 1 because it goes counterclockwise around 0 once as θ goes from 0 to 2π. For the lead operator $\mathbf{F}$, the symbol is z^{-1}, which has graph $\cos\theta - \sin\theta \cdot i$ and a winding number of -1, because it goes clockwise around 0 once. Hence $\mathbf{L} : (x_0, x_1, \ldots) \mapsto (0, x_0, x_1, \ldots)$ has less than full range, $\text{nonex}(\mathbf{L}) = 1$, and $\mathbf{F} : (x_0, x_1, \ldots) \mapsto (x_1, x_2, \ldots)$ has a one-dimensional null space, $\text{indet}(\mathbf{F}) = 1$.

We may also use this criterion to check the properties of the neoclassical growth model in the previous section. $\mathbf{H_K}$'s symbol, from equation (22), has graph

$$\left((1 + f_K)(\cos\theta - 1) - \frac{f_{KK}}{f_K}\frac{u'}{u''}\right) + (f_K - 1)\sin\theta \cdot i.$$

If both f and u are increasing and concave at the steady state, the graph lies entirely to the left of the imaginary axis. It will never wind around the origin, so its winding number will be zero. As a result, $\mathbf{H_K}$ is invertible and we have a unique rational expectations solution. This is illustrated by the green line in figure 6.

We also consider a modification of this model, detailed more fully in appendix B.9, where production externalities given by γ increase the returns to scale, as in Baxter and King (1991) and Benhabib and Farmer (1994). For a high enough γ, this model suffers from local *nonexistence*, as revealed by the winding number of 1 in figure 6 for the gray dashed line, which winds once counterclockwise around the origin. This is because the model's dynamics are locally explosive, making it impossible to solve for a bounded im-

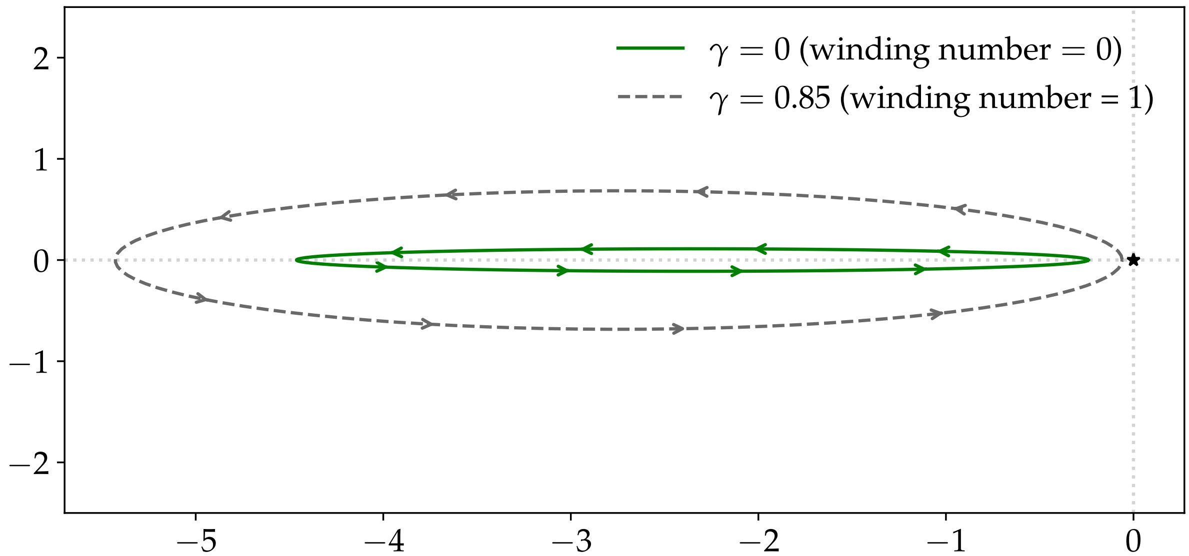


Figure 6: Determinacy of neoclassical growth model by production externality γ

pulse response to most shocks.[19]

Relation to standard criteria for linear difference equations. Consider a version of (19) with more leads and lags

$$b_0 u_t + \sum_{j=-m}^{n} a_j \mathbb{E}_t[x_{t-j}] = 0. \tag{24}$$

A standard approach to assessing existence and uniqueness for (24) would be to write the characteristic polynomial $\widehat{a}(\lambda) \equiv a_n + a_{n-1}\lambda + \ldots + a_{-m}\lambda^{m+n}$ of the implied recurrence for x_t in the absence of shocks. Following Blanchard and Kahn (1980), there is a unique solution to (24) if $\widehat{a}(\lambda)$ has n stable roots ($|\lambda| < 1$). If it has fewer stable roots, there is nonexistence; if it has more, there is indeterminacy.

Alternatively, we can write (24) in the sequence space as $T(\mathbf{a})\mathbf{x} = -b_0\mathbf{z}$, and apply proposition 6. To connect to the root-counting test above, we will use the fact that for any rational function j, the winding number $\text{wind}(j)$ equals the number of roots inside the unit circle minus the number of poles, counted with multiplicity.[20]

The symbol of $T(\mathbf{a})$ is $a(z) = a_{-m}z^{-m} + \ldots + a_n z^n$. We note that $a(z^{-1})$ has minus

[19]In models with endogenous labor, such as Benhabib and Farmer (1994), increasing γ from 0 often produces indeterminacy instead.

[20]This fact actually holds whenever j is a meromorphic function on the unit disk (as guaranteed by j being rational). This can be derived by substituting $z = \gamma(w)$ into (23) and applying the argument principle.

the winding number of $a(z)$, since z^{-1} goes the opposite direction of z on the unit circle. Further, $a(z^{-1})$ equals $z^{-n}\widehat{a}(z)$, whose winding number equals the number of stable roots of $\widehat{a}$, minus n (because z^{-n} has a pole at $z = 0$ of multiplicity n). Hence, $\mathrm{wind}(a)$ equals zero if $\widehat{a}$ has exactly n stable roots—consistent with the Blanchard-Kahn criterion above for a unique solution. $\mathrm{wind}(a)$ is positive if there are fewer than n stable roots, and vice versa—also consistent with the criteria for nonexistence and indeterminacy above.

In short, the Blanchard-Kahn and winding number tests agree, as earlier shown by Onatski (2006). The benefit of the winding number test is that it is broader, since it applies even when (24) has infinitely many leads and lags—and also, as we now show, to more general systems.

3.2 The quasi-Toeplitz case and genericity

As we saw earlier, the sequence-space Jacobians of heterogeneous-agent models are generally quasi-Toeplitz rather than Toeplitz. To assess whether there is a unique solution in such cases—for instance, for the intertemporal Keynesian cross equation $\mathbf{A}(d\mathbf{Y} - d\mathbf{T}) = d\mathbf{B}$ with quasi-Toeplitz $\mathbf{A}$—we must extend the winding number criterion.

Unlike in proposition 6, in the quasi-Toeplitz case it is possible that $\mathbf{Jx} = \mathbf{y}$ suffers from both nonexistence and multiplicity. The following weaker version of the proposition, however, still holds.

Proposition 7. *For any quasi-Toeplitz operator* $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ *where* $\mathrm{wind}(j)$ *is defined, both* $\mathrm{nonex}(\mathbf{J})$ *and* $\mathrm{indet}(\mathbf{J})$ *are finite, and*

$$\mathrm{wind}(j) = \mathrm{nonex}(\mathbf{J}) - \mathrm{indet}(\mathbf{J}). \tag{25}$$

Proof. See appendix C.2. ■

Now, even if the winding number is zero, (25) leaves open the possibility that $\mathrm{nonex}(\mathbf{J})$ and $\mathrm{indet}(\mathbf{J})$ are equal and positive, in which case $\mathbf{J}$ is not invertible. Importantly, however, if $\mathbf{J}$ is invertible, then the winding number *must* be zero. In other words, if we use $\mathrm{wind}(j) = 0$ as a diagnostic for the existence of a unique solution, there may be false positives (winding number is zero where there is not a unique solution), but never false negatives (winding number is non-zero when a unique solution exists).

Ruling out false positives. There are two pieces of very good news about false positives. First, nonexistence and multiplicity must go hand-in-hand. If the winding number is zero but we suspect multiplicity, then we can test for nonexistence through other means; if we

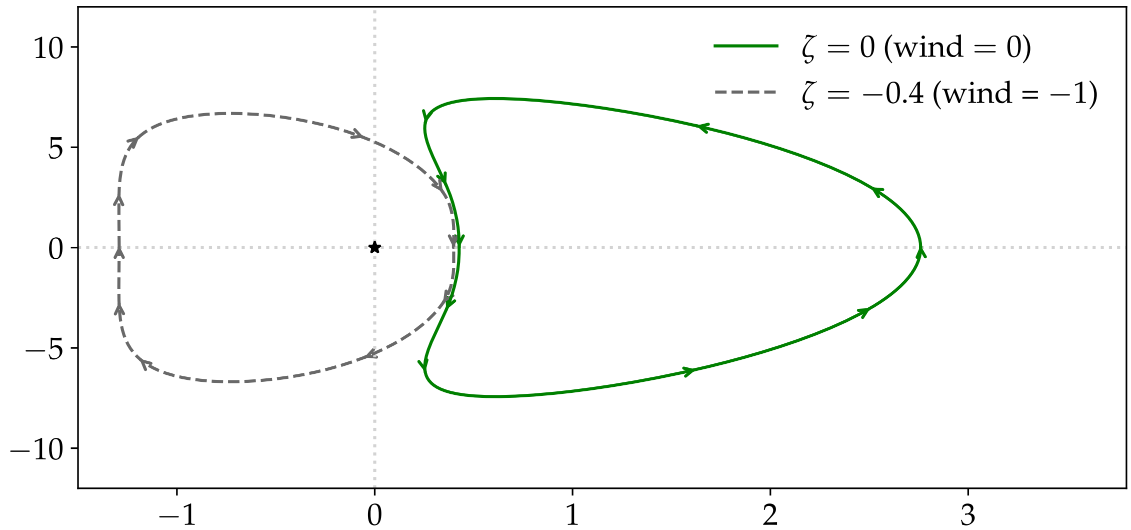


Figure 7: Winding number of asset Jacobian symbol $a(z)$

rule out nonexistence, we then also rule out multiplicity. Section 3.5 discusses how to do this in practice. Second, proposition 6 holds *generically*, in the following sense.

Proposition 8. *Proposition 6 holds for* generic *quasi-Toeplitz* $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$*, i.e. on an open and dense subset (in the operator norm) of all quasi-Toeplitz operators.*

Proof. See appendix C.3. ∎

For any quasi-Toeplitz $\mathbf{J}$, therefore, either proposition 6 holds, or it holds for some arbitrarily close $\tilde{\mathbf{J}}$. If it holds for $\mathbf{J}$, it also holds in a neighborhood of $\mathbf{J}$.

Topologically, an open and dense subset fills almost all of a space, in the sense that its complement has empty interior. The use of "generic" to mean "on an open and dense subset" is common both in mathematics and economics; for examples of the latter, see Hurwicz and Walker (1990) and Weinstein and Yildiz (2007).

In economic applications, we have found that non-genericity is rare. For a simple example, however, we can consider $\mathbf{LF}$. From (7), we see that it is quasi-Toeplitz, equal to the identity $\mathbf{I}$ plus a compact correction $\mathbf{E}$ that has all zeros except for the $(0,0)$ entry, which is equal to -1. Its winding number is the same as $\mathbf{I}$, zero, but it is clearly not invertible. But taking, for instance, $\mathbf{LF} + \varepsilon\mathbf{E}$ for any $\varepsilon \neq 0$, including arbitrarily small ε, gives us an invertible operator, in line with the winding number of zero.

Summary of results. We summarize our results for quasi-Toeplitz operators in the following proposition.

Proposition 9. *Consider a quasi-Toeplitz operator* $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ *where* $j(z) \neq 0$ *for* $z \in \mathbb{T}$*. Then:*

1. *If* $\text{wind}(j) = 0$*, generically* $\mathbf{J}$ *is invertible, and if the inverse* $\mathbf{J}^{-1}$ *exists it is quasi-Toeplitz.*

2. *If* $\text{wind}(j) < 0$*, any solution to* $\mathbf{Jx} = \mathbf{y}$ *is not unique,* $\text{indet}(\mathbf{J}) > 0$*, but generically a solution always exists,* $\text{nonex}(\mathbf{J}) = 0$*.*

3. *If* $\text{wind}(j) > 0$*, there are* $\mathbf{y}$ *for which a solution to* $\mathbf{Jx} = \mathbf{y}$ *does not exist,* $\text{nonex}(\mathbf{J}) > 0$*, but generically all solutions that do exist are unique,* $\text{indet}(\mathbf{J}) = 0$*.*

In the non-generic case, there is both non-uniqueness and non-existence: $\text{indet}(\mathbf{J}), \text{nonex}(\mathbf{J}) > 0$*.*

Proof. See appendix C.4. ∎

Generically, a quasi-Toeplitz $\mathbf{J}$ is invertible if and only if its Toeplitz part $T(\mathbf{j})$ is invertible, since the same winding number test applies for both. It is worth emphasizing that the class of quasi-Toeplitz operators is closed under *inversion* as well as multiplication, so long as an inverse exists.

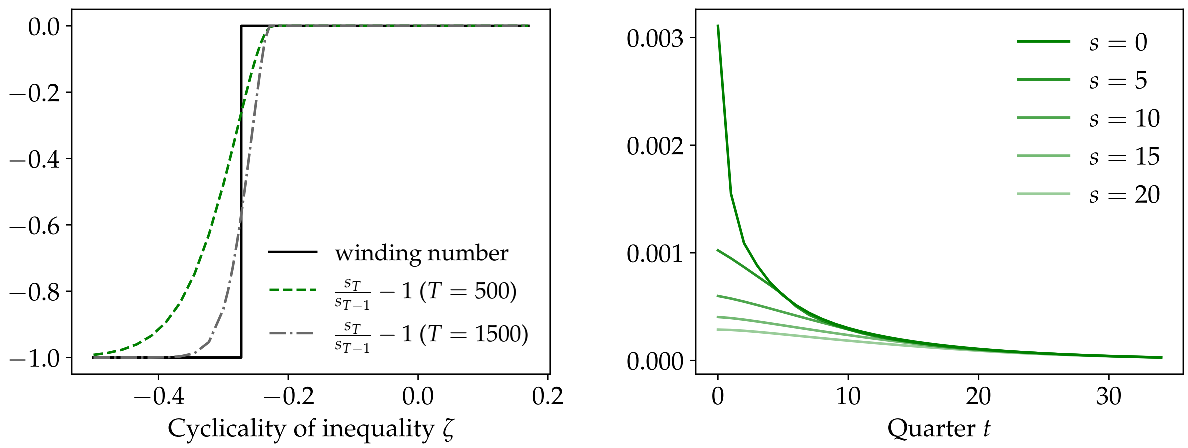
3.3 Application to the intertemporal Keynesian cross

We now apply our results to see whether the intertemporal Keynesian cross in asset space, $\mathbf{A}(d\mathbf{Y} - d\mathbf{T}) = d\mathbf{B}$, has a unique solution. To do so, we calculate the winding number of the symbol $a(z)$ of $\mathbf{A}$, to see whether $\mathbf{A}$ is invertible. We do so both for the main calibration of the model, for which we plotted $\mathbf{A}$ in figure 3(b), and also an alternative calibration discussed in appendix B.6 where the incidence of aggregate income changes is higher on the poor than the rich. This latter case makes idiosyncratic income risk and inequality countercyclical, since the gap between poor and rich grows in recessions.

Figure 7 plots the graphs of $a(z)$, when evaluated around the unit circle $\mathbb{T}$, in the complex plane for both cases. We see that there is a winding number of 0 for the main case, since the graph does not wrap around the origin $(0, 0)$, but a winding number of -1 for the countercyclical inequality case, since the graph wraps clockwise around the origin. This echoes similar findings of indeterminacy in analytical versions of HANK models with countercyclical risk (e.g. Ravn and Sterk 2021, Acharya and Dogra 2020, and Bilbiie 2024.)

Figure 8(a) plots the winding number for a range of calibrations, where a negative ζ corresponds to countercyclical risk.[21] The winding-number test gives a precise answer for the level of ζ at which the model shifts from indeterminacy to determinacy: $\zeta = -0.272$.

[21] In our parametrization, explained in appendix B.6, ζ is the sensitivity of the cross-sectional standard deviation of labor income to log changes in aggregate labor.



(a) Winding number and SVD ratio tests

(b) Columns of compact correction for $\mathbf{A}^{-1}$

Figure 8: Winding number test and quasi-Toeplitz structure of $\mathbf{A}^{-1}$

An alternative test is to plot the ratio of the smallest and second-smallest singular values of a $T \times T$ truncation of the Toeplitz part $T(\mathbf{a})$ of $\mathbf{A}$. When $T(\mathbf{a})$ is not invertible, the smallest singular value s_T should be close to zero, while the second-smallest s_{T-1} is not—and the gap should become more pronounced as T increases.[22] When there is determinacy, on the other hand, the ratio should be close to 1. Figure 8(a) compares $s_T/s_{T-1} - 1$ to the winding number for both $T = 500$ and $T = 1500$. Especially for higher T, this ratio test roughly lines up with the winding number, but it is less precise and also computationally much more costly.

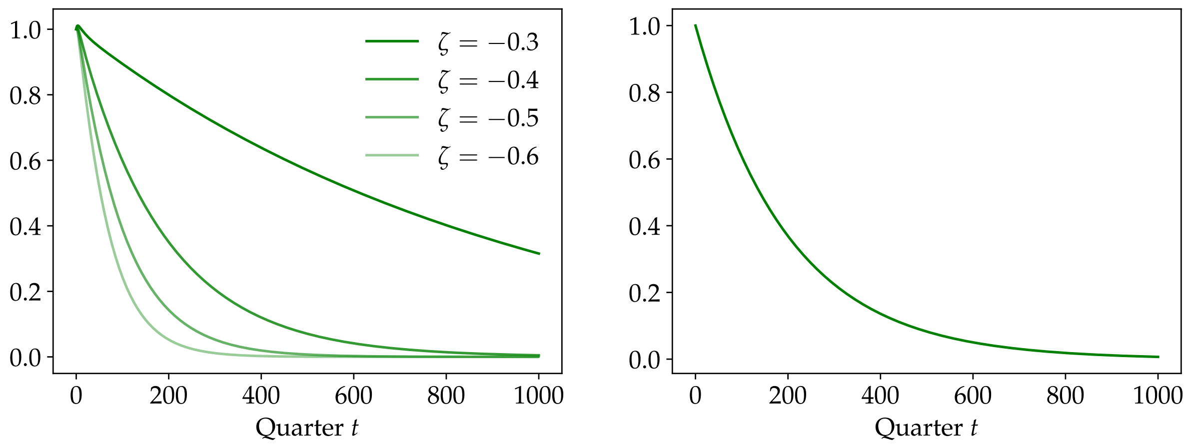
Finally, figure 8(b) plots columns of the compact correction $\mathbf{A}^{-1} - T(\mathbf{a}^{-1})$ of $\mathbf{A}^{-1}$ for our baseline calibration with $\zeta = 0$. We see that not only is $\mathbf{A}^{-1}$ quasi-Toeplitz, but the distance from Toeplitz is also quite small.

3.4 Characterizing indeterminacy or nonexistence

Suppose that the winding number test tells us that our quasi-Toeplitz system $\mathbf{J}\mathbf{x} = \mathbf{y}$ suffers from indeterminacy or nonexistence. It is natural to ask what this indeterminacy or nonexistence looks like: for what $\mathbf{x}$ do we have $\mathbf{J}\mathbf{x} = 0$, and for what $\mathbf{y}$ is there no solution?

Figure 9(a) plots the "self-fulfilling" output changes $d\mathbf{Y}$ that satisfy $\mathbf{A}d\mathbf{Y} = 0$, for various levels of ζ where there is indeterminacy, normalizing dY_0 to 1. In the language of

[22] Section 4.6 of Böttcher and Silbermann (2012) puts this on formal footing, showing that as $T \to \infty$, the set of singular values of the truncated matrix approaches the true set of singular values for the Toeplitz operator. We can also perform this "SVD ratio" test directly on the quasi-Toeplitz $\mathbf{A}$, and the results are similar but somewhat less accurate for low T.



(a) Self-fulfilling output $d\mathbf{Y} \in \text{null } \mathbf{A}$

(b) Missing $\mathbf{x} \in (\text{range } \mathbf{I} - \mathbf{M})^{\perp}$

Figure 9: Indeterminacy and nonexistence examples in the IKC model

the goods-market Intertemporal Keynesian Cross, these $d\mathbf{Y}$ satisfy $d\mathbf{Y} = \mathbf{M}d\mathbf{Y}$: they are increases in output that generate an exactly equal amount of consumption demand.

Here, these $d\mathbf{Y}$ are roughly exponential: higher future dY_{t+h}, by lowering anticipated idiosyncratic risk, drives an even larger increase in spending dY_t at date t. When we are close to the indeterminacy threshold, as with $\zeta = -0.3$, the path is *extremely* persistent: to sustain an increase in output today, we need an anticipated increase in output 30% as large, even 1000 quarters in the future. Indeed, if we keep increasing ζ to the threshold where the model becomes determinate, persistence approaches 1, as an increase in spending needs to be almost permanent to be self-fulfilling. For more negative ζ, on the other hand, the path features far less persistence.

Figure 9(b) tackles a case with nonexistence. If we rearrange the goods-market intertemporal Keynesian cross (briefly discussed in section 2.5) to isolate $d\mathbf{Y}$, we get $(\mathbf{I} - \mathbf{M})d\mathbf{Y} = d\mathbf{G} - \mathbf{M}d\mathbf{T}$. It is tempting to solve this by inverting $\mathbf{I} - \mathbf{M}$, but as Auclert et al. (2024a) point out, $\mathbf{I} - \mathbf{M}$ is not invertible, because its range only includes sequences with net present value zero. Applying the winding number test confirms this: the symbol $1 - m(z)$ for our baseline calibration has a winding number of 1. We can visualize the nonexistence here by finding the sequence to which the range of $\mathbf{I} - \mathbf{M}$ is orthogonal. This is simply $\mathbf{q} \equiv \{1, (1+r)^{-1}, (1+r)^{-2}, \ldots\}$, as plotted in figure 9(b). For a solution to the goods-market intertemporal Keynesian cross to exist, the right side $d\mathbf{G} - \mathbf{M}d\mathbf{T}$ must be orthogonal to $d\mathbf{Y}^{nonex}$, i.e. it must have net present value zero. Fortunately, this is guaranteed whenever government debt does not explode.[23]

[23]Interestingly, if we premultiply both sides of the goods-market intertemporal Keynesian cross by $\mathbf{F}$, i.e.

In figure 9(a), we obtain self-fulfilling output by numerically looking for a sequence that is in null $\mathbf{A}$. Concretely, we find the right singular vector for the smallest singular value of a truncated $\mathbf{A}$.[24] Similarly, in figure 9(b), we look for a sequence that is in $(\text{range } \mathbf{I} - \mathbf{M})^{\perp}$ by finding the left singular vector for the smallest singular value of a truncated $\mathbf{I} - \mathbf{M}$.

3.5 Non-genericity

In proposition 9, the winding number test applies *generically* for quasi-Toeplitz operators, meaning that it gives the correct answer on an open and dense subset of all such operators.

Topologically, non-generic operators are rare, but it is possible that the economic structure of a problem makes it non-generic. In other words, it is important to distinguish between genericity in the space of *operators*, which is relevant for proposition 9, and genericity in the space of *economic models of interest*. Even if the winding number test applies generically in the former space, it might not hold generically in the latter.

We know of one main example, which is the following. Suppose that in the previous section, we are in the case with indeterminacy, where the winding number of $a(z)$ is -1. In this case, it is possible to show that the winding number of $1 - m(z)$ is 0.[25] This would suggest that $\mathbf{I} - \mathbf{M}$ is invertible, but its range still has only sequences with net present value zero, so it is clearly not invertible. This is a *non-generic* case where the winding number test fails.

Suppose we were not aware of this issue, and naively concluded from the winding number test that $\mathbf{I} - \mathbf{M}$ was invertible when $\zeta = -0.4$. How could we detect our mistake? As proposition 9 shows, in the non-generic case, we always have both indeterminacy and nonexistence. We can test for the latter numerically by feeding in arbitrary $\mathbf{y}$ that decays to zero and seeing if $(\mathbf{I} - \mathbf{M})\mathbf{x} = \mathbf{y}$ has any solution—which it will not unless $\mathbf{y}$ happens to lie in the right subspace. To implement this, we choose 100 sequences $y_t \equiv \rho_1^t - 0.5 \cdot \rho_2^t$, with random draws of $\rho_1, \rho_2 \in [0.85, 0.95]$ for each sequence. Solving the truncated system

we delete the first row, then $\mathbf{F}(\mathbf{I} - \mathbf{M})$ has a winding number of 0 (this is the sum of the winding number of $\mathbf{F}$, which is -1, and the winding number of $\mathbf{I} - \mathbf{M}$), and is invertible. Indeed, as the appendix of Auclert et al. (2024a) mentions, deleting the first row is an effective way of solving the IKC in goods space.

[24] Interestingly, if we ask for the *left* singular vector for the smallest singular value of a truncated $\mathbf{A}$, we get an explosive sequence that is largest near the truncation boundary T. This is because the true, infinite-dimensional $\mathbf{A}$ has full range, so if anything is missing from the truncated $\mathbf{A}$'s range, it must be closely tied to truncation.

[25] This is an instance of the general fact that $\text{wind}(1 - m) = \text{wind}(a) + 1$. To see this, note that from the aggregated budget constraint, $\mathbf{I} - \mathbf{M} = (\mathbf{I} - (1 + r)\mathbf{L})\mathbf{A}$, so we have $1 - m(z) = (1 - (1 + r)z)a(z)$. Since the winding number of a product of two symbols equals the sum of the winding numbers, $\text{wind}(1 - m)$ equals $\text{wind}(a)$ plus the winding number of $1 - (1 + r)z$. The latter is 1 because there is a single zero inside the unit circle at $z = 1/(1 + r)$ and no poles.

with $T = 1000$, we obtain $\mathbf{x}$ with entries of at least 2000 on every draw, and over 200,000 on some draws—many orders of magnitude larger than the $\mathbf{y}$, and economically unrealistic. This indicates that there is not a true solution to $(\mathbf{I} - \mathbf{M})\mathbf{x} = \mathbf{y}$ for most $\mathbf{y}$, revealing that $\mathbf{I} - \mathbf{M}$ is not invertible.[26]

We can always sharpen this analysis with a higher T; for instance, with $T = 2000$, the same exercise finds "solutions" $\mathbf{x}$ with entries of almost 1 billion.[27]

How can it be that $\mathbf{I} - \mathbf{M}$ is non-generic, lying outside an open and dense set that covers almost the entire space of operators? The reason is that the present-value household budget constraint restricts $\mathbf{M}$ to have a very special structure, where if $\mathbf{q} \equiv (1, (1+r)^{-1}, (1+r)^{-2}, \ldots$, then $\mathbf{q}'(\mathbf{I} - \mathbf{M}) = 0$. Nearly all "nearby" $\mathbf{M}$, in the sense of the operator norm, are generic, but these do not satisfy $\mathbf{q}'(\mathbf{I} - \mathbf{M}) = 0$ and are therefore not economically valid. This echoes the observation by Sims (2007) that theoretically motivated constraints may make a model non-generic.

In practice, we have found that solving models in the goods space, where such intertemporal budget constraints are present, can lead to non-genericity in other cases as well, such as the quantitative IKC model in appendix D.2. We have not, however, found any practical example of non-genericity when a model is solved in the asset space. It is therefore best to solve at least partly in the asset space—imposing asset market clearing, and obtaining goods market clearing via Walras's law.[28]

Irrespective of the approach taken to solve the model, the key fact about non-genericity is that it implies *both* indeterminacy and nonexistence. This allows us to test indeterminacy, if we are worried about non-genericity, by asking the easier question of nonexistence.

3.6 Application to Taylor principle with heterogeneous agents

We now extend the model of section 3.3 to allow for a time-varying real interest rate, consistent with a New Keynesian Phillips curve for inflation and a Taylor rule for the nominal interest rate. To do so, we need to set up a larger sequence-space system.

For this example, we assume that the Phillips curve takes the form $\pi_t = \kappa dY_t + \beta \mathbb{E}\pi_{t+1}$, which we can write in sequence space as $\boldsymbol{\pi} = \mathbf{K} d\mathbf{Y}$, where $\mathbf{K}$ is upper-triangular and has entries $K_{t,s} = \beta^{s-t}\kappa$ for $s \geq t$. We also assume that the Taylor rule is $i_t = r + \phi\pi_t$, where r is the steady-state real interest rate, and therefore $dr_t = \phi\pi_t - \mathbb{E}\pi_{t+1}$, which can be

[26] Since finite-dimensional matrices are generically invertible, the truncated version of $\mathbf{I} - \mathbf{M}$ usually will be invertible even if the infinite-dimensional $\mathbf{I} - \mathbf{M}$ operator features non-existence. This makes this type of procedure necessary.

[27] As in section 3.3, we can also look at the smallest singular value and see if it is close to zero.

[28] As Auclert and Rognlie (2020) point out, it is also dangerous to solve in goods space in the steady state, since goods-market clearing can be satisfied without asset market clearing when $r = 0$.

written in sequence space as $d\mathbf{r} = (\phi\mathbf{I} - \mathbf{F})\boldsymbol{\pi}$. Finally, as in Auclert, Rognlie and Straub (2025), we assume that the fiscal rule adjusts taxes T_t to target constant debt inclusive of interest $(1 + r_t)B_t$ going into the next period. In the absence of any shocks to government spending, this implies that $(1 + r)dB_t = -Bdr_t$, and $dT_t = -dB_t$.

Defining $\mathbf{A}^r$ to be the Jacobian of aggregate assets with respect to the real interest rate, asset-market clearing implies that $\mathbf{A}(d\mathbf{Y} - d\mathbf{T}) + \mathbf{A}^r d\mathbf{r} = d\mathbf{B}$, which using $dT_t = -dB_t$ can be rewritten as $\mathbf{A}d\mathbf{Y} + \mathbf{A}^r d\mathbf{r} = (\mathbf{I} - \mathbf{A})d\mathbf{B}$. Finally, using $dB_t = -(1 + r)^{-1}Bdr_t$, this becomes

$$
\begin{aligned}
0 &= \mathbf{A}d\mathbf{Y} + \left(\mathbf{A}^r + (1+r)^{-1}B(\mathbf{I} - \mathbf{A})\right) d\mathbf{r} \\
&= \left(\mathbf{A} + \left(\mathbf{A}^r + (1+r)^{-1}B(\mathbf{I} - \mathbf{A})\right)(\phi\mathbf{I} - \mathbf{F})\mathbf{K}\right) d\mathbf{Y}
\end{aligned} \tag{26}
$$

where in the second line we substitute $d\mathbf{r} = (\phi\mathbf{I} - \mathbf{F})\boldsymbol{\pi} = (\phi\mathbf{I} - \mathbf{F})\mathbf{K}d\mathbf{Y}$.

Winding number test and determinacy. Our question now is determinacy: for a given Taylor rule coefficient ϕ, is $d\mathbf{Y} = 0$ the only solution to (26)?

To do so, we apply the winding number test to the operator multiplying $d\mathbf{Y}$ in (26), which we denote by $\boldsymbol{\Theta}$. This has symbol

$$
\theta(z) = a(z) + \left(a^r(z) + (1+r)^{-1}B(1 - a(z))\right)(\phi - z^{-1})\kappa(1 - \beta z^{-1})^{-1} \tag{27}
$$

where $a(z)$ and $a^r(z)$ are the symbols of $\mathbf{A}$ and $\mathbf{A}^r$, z^{-1} is the symbol of $\mathbf{F}$, and we have substituted $k(z) = \kappa(1 + \beta z^{-1} + \ldots) = \kappa(1 - \beta z^{-1})^{-1}$ for the symbol of $\mathbf{K}$. In going from (26) to (27), we use the important fact from proposition 1 that the symbol of a product is the product of the symbols.

As we vary the Taylor coefficient ϕ or the slope κ of the Phillips curve, the only part of (27) that changes is the factor $(\phi - z^{-1})\kappa$ in the second term. As a result, the winding number test can be implemented extremely efficiently across different ϕ and κ, since we can precalculate everything except $(\phi - z^{-1})\kappa$ for our sample points z around the unit circle. Figure 10 shows the results in (ϕ, κ) space.

If the Phillips curve has slope $\kappa = 0$, then ϕ is irrelevant and we always have determinacy; in this case, $\theta(z) = a(z)$ in (27) and we are back at our original case of a constant real rate rule. For high κ, on the other hand, the threshold for determinacy approaches the traditional Taylor principle $\phi = 1$.[29] In between, the Taylor rule threshold ϕ is strictly lower

[29] This can be seen from (27) as follows. As $\kappa \to \infty$, the right term dominates. One can numerically calculate that the winding number of $\left(a^r(z) + (1+r)^{-1}B(1 - a(z))\right)(1 - \beta z^{-1})^{-1}$ is zero. In the limit, therefore,

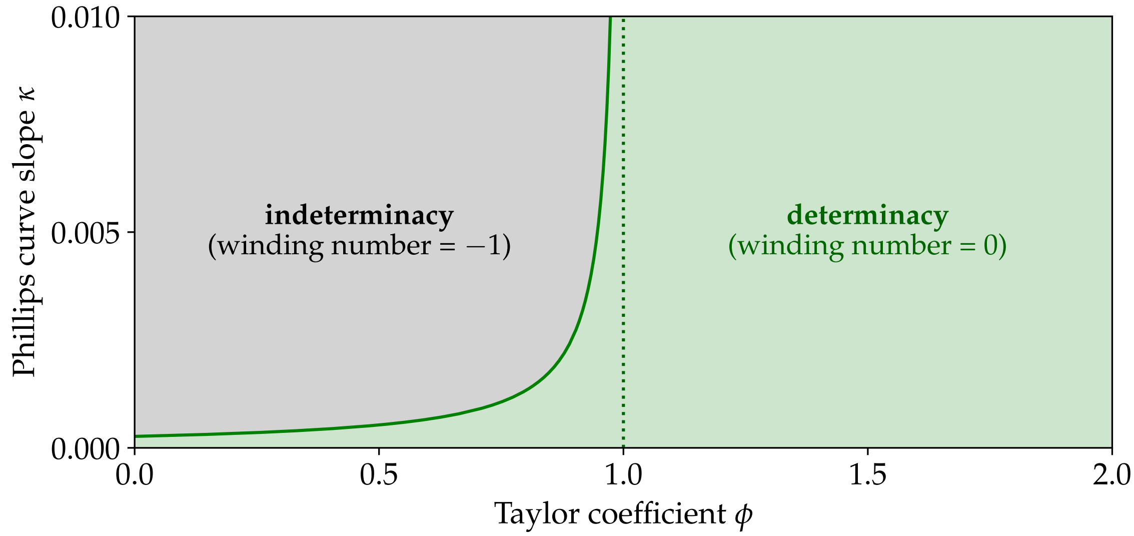


Figure 10: Regions of indeterminacy and determinacy for (26) in (ϕ, κ) space

than 1, though often only by a small amount. For instance, at $\kappa = 0.0062$ (the quarterly Phillips curve slope with respect to unemployment from Hazell, Herreño, Nakamura and Steinsson 2022), the determinacy threshold in figure 10 is $\phi = 0.96$.

3.7 Extending to the block case

Up to this point we have considered *scalar* systems, where we solve $\mathbf{J}\mathbf{x} = \mathbf{y}$ for a single unknown sequence $\mathbf{x} \in \ell^2$, given a single target sequence $\mathbf{y} \in \ell^2$. In practice, however, we often need to solve for multiple unknown sequences to hit an equal number n of targets. We can represent such cases with a *block* system.

In a block system $\mathbf{J}\mathbf{x} = \mathbf{y}$, $\mathbf{x}$ and $\mathbf{y}$ are sequences of vectors rather than scalars, and the Jacobian $\mathbf{J}$ is a map between sequences of vectors. Formally, we say that $\mathbf{x}$ and $\mathbf{y}$ are in the space $\ell^2(\mathbb{R}^n)$, rather than the previous $\ell^2(\mathbb{R})$, and $\mathbf{J}$ is an operator on $\ell^2(\mathbb{R}^n)$. An entry $J_{t,s}$ of $\mathbf{J}$ is now an $n \times n$ matrix, representing the effect of the vector x_s on the vector y_t.

Very little changes in the process of accommodating the block case. Definitions 1–3 for Toeplitz operators, quasi-Toeplitz operators, and their symbols immediately generalize.

Definition 8. *We say a linear operator* $\mathbf{J}$ *on* $\ell^2(\mathbb{R}^n)$ *is* block Toeplitz $T(\mathbf{j})$ *if its matrix has the form* (4) *with constant entries* j_k *on each diagonal, where the* j_k *are matrices on* $\mathbb{R}^n$ *and are absolutely summable (i.e.* $\sum_{k=-\infty}^{\infty} |j_k| < \infty$*, with* $|j_k|$ *being the matrix norm of* j_k*). The* symbol *of* $T(\mathbf{j})$ *is the function* $j : \mathbb{T} \to \mathbb{C}^{n\times n}$ *given by* $j(z) \equiv \sum_{k=-\infty}^{\infty} j_k z^k$*.*

wind θ equals the winding number of $\phi - z^{-1}$, which is 0 for $\phi > 1$ and -1 for $\phi < 1$.

We say that $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ *is* block quasi-Toeplitz *if* $T(\mathbf{j})$ *is block Toeplitz and* $\mathbf{E}$ *is a compact operator on* $\ell^2(\mathbb{R}^n)$.

Proposition 10. *Lemma 1 and propositions 1–3 hold with "Toeplitz" and "quasi-Toeplitz" replaced by "block Toeplitz" and "block quasi-Toeplitz".*

Proof. See appendix C.5 ∎

We can think of a block Toeplitz operator $T(\mathbf{j})$ on $\ell^2(\mathbb{R}^n)$ as consisting of $n \times n$ scalar Toeplitz operators $T(\mathbf{j}_{u,v})$, each of which maps sequences in some dimension u to sequences in some dimension v of $\mathbb{R}^n$. The same is true for block quasi-Toeplitz.

Winding number criterion. The winding number criterion for determinacy and existence also generalizes to the block case, with two important details. First, since the symbol is now matrix-valued, we take the winding number of its *determinant*.

Second, the exact winding number criterion for Toeplitz operators in proposition 6 does not extend to the block Toeplitz case. Instead, with blocks, we have the same criterion as for quasi-Toeplitz.

Proposition 11. *Propositions 7–9 hold for "quasi-Toeplitz" replaced by "block quasi-Toeplitz", and* $\text{wind}(j)$ *replaced by* $\text{wind}(\det j)$.

Proof. See appendix C.5. ∎

Thus the winding number test holds for *generic* block quasi-Toeplitz $\mathbf{J}$ (as previously derived by Onatski 2006 for the block Toeplitz case), and in non-generic cases, we have both indeterminacy and nonexistence.

Appendix B.8 shows that for multidimensional linear difference equations (21), the winding number test is equivalent to the standard root-counting criterion for existence and uniqueness.

Application to Taylor principle. In the last section, we used substitutions to reduce our model to a single, somewhat complicated scalar Toeplitz system (26) in $d\mathbf{Y}$. Alternatively, we can avoid these substitutions and write the system in $[d\mathbf{Y}, d\mathbf{r}, \boldsymbol{\pi}]$ as

$$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{A}^r + (1+r)^{-1}B(\mathbf{I} - \mathbf{A}) & 0 \\ 0 & \mathbf{I} & -(\phi\mathbf{I} - \mathbf{F}) \\ -\mathbf{K} & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} d\mathbf{Y} \\ d\mathbf{r} \\ \boldsymbol{\pi} \end{bmatrix} \tag{28}$$

The block symbol corresponding to the matrix in (28) is

$$\begin{bmatrix} a(z) & a^r(z) + (1+r)^{-1}B(1-a(z)) & 0 \\ 0 & 1 & -(\phi - z^{-1}) \\ -k(z) & 0 & 1 \end{bmatrix} \tag{29}$$

and we can then apply the winding number test to the determinant of (29). This delivers the same results as in the previous section.[30]

4 Improving on truncation

4.1 Truncation and spurious missing anticipation

The most common way to use sequence-space Jacobians in practice is to truncate them and work with the resulting finite-dimensional matrices. For instance, we replace the infinite-dimensional system $\mathbf{J}\mathbf{x} = \mathbf{y}$ with a finite-dimensional version $\mathbf{J}_{T,T}\mathbf{x}_T = \mathbf{y}_T$. In the mathematical literature, this is called the *finite section* method, and there are results (Böttcher and Silbermann 2012), summarized in the following proposition, showing convergence to the true solution as $T \to \infty$.

Proposition 12. *For a system $\mathbf{J}\mathbf{x} = \mathbf{y}$ with $\mathbf{x}, \mathbf{y} \in \ell^2(\mathbb{R}^n)$, define $\mathbf{J}_{T,T}$ to be the truncated $T \times T$ submatrix of $\mathbf{J}$, $\mathbf{y}_T$ to be the length-T truncation of $\mathbf{y}$, and $\widehat{\mathbf{x}}_T \equiv \mathbf{J}_T^{-1}\mathbf{y}_T$ to be the solution of the truncated system. If $\mathbf{J}$ is quasi-Toeplitz ($n = 1$), then $\widehat{\mathbf{x}}_T$ converges to the true solution $\mathbf{x} = \mathbf{J}^{-1}\mathbf{y}$ as $T \to \infty$, assuming $\mathbf{J}$ is invertible. This is generically also true when $\mathbf{J}$ is block quasi-Toeplitz ($n > 1$).*

Proof. In appendix D.1. ■

In short, if the winding number test from section 3 tells us that $\mathbf{J}\mathbf{x} = \mathbf{y}$ always has a unique solution, so that $\mathbf{J}$ is invertible, then for large enough T, the truncated system will get arbitrarily close to the correct answer. This is encouraging, and suggests that truncation is a reasonable approach.

Spurious missing anticipation. Still, for any finite T, truncation introduces distortions. Conceptually, these distortions are the mirror image of the "missing anticipation" discussed in section 2. By truncating the system, we assume that all agents in the model

[30] Indeed, in this case, it is easy to show analytically that the determinant of (29) is exactly $\theta(z)$ from (27).

expect x_t and y_t to be zero for $t \geq T$. If the agents are forward-looking, the $\widehat{\mathbf{x}}_T$ we compute is incorrectly missing the anticipation of both the shock itself and any endogenous propagation. Unlike the missing anticipation discussed earlier—which was part of the economic environment, with shocks only realized at $t = 0$—this missing anticipation is purely an artifact of truncating the system, and we call it *spurious missing anticipation*.

For a simple example, consider $\mathbf{FL}$, the product of the lead and lag operators. Without truncation, this is the identity ($\mathbf{FL} = \mathbf{I}$), but the truncated product is $\mathbf{F}_T\mathbf{L}_T \neq \mathbf{I}_T$:

$$\begin{pmatrix} \ddots & \ddots & \ddots & \\ \ddots & 0 & 1 & 0 \\ \ddots & 0 & 0 & 1 \\ & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ddots & \ddots & \ddots & \\ \ddots & 0 & 0 & 0 \\ \ddots & 1 & 0 & 0 \\ & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \ddots & \ddots & \ddots & \\ \ddots & 1 & 0 & 0 \\ \ddots & 0 & 1 & 0 \\ & 0 & 0 & 0 \end{pmatrix}. \tag{30}$$

Here, the final entry on the diagonal is a zero instead of a one. This is reminiscent of $\mathbf{LF}$ in (7), but with the missing entry in the opposite corner of the matrix. This indicates two important properties of spurious missing anticipation: it results from taking leads of lags, and it is confined to be near the truncation boundary at (T, T), just as the compact correction is near $(0, 0)$.

Decomposing the missing anticipation. To better understand the error introduced by spurious missing anticipation, we can partition the infinite system into blocks,

$$\mathbf{Jx} = \begin{bmatrix} \mathbf{J}_{T,T} & \mathbf{J}_{T,+} \\ \mathbf{J}_{+,T} & \mathbf{J}_{+,+} \end{bmatrix} \begin{bmatrix} \mathbf{x}_T \\ \mathbf{x}_+ \end{bmatrix} = \begin{bmatrix} \mathbf{y}_T \\ \mathbf{y}_+ \end{bmatrix} = \mathbf{y}, \tag{31}$$

where as before the subscript T denotes truncation to length T, and now $+$ denotes the remaining part from T to ∞. We can then write the error in the truncated solution $\widehat{\mathbf{x}}_T = \mathbf{J}_{T,T}^{-1}\mathbf{y}_T$ as the sum of "endogenous" error ϵ_T^n and "exogenous" error ϵ_T^x,

$$\widehat{\mathbf{x}}_T - \mathbf{x}_T = \epsilon_T^n + \epsilon_T^x = \epsilon_T, \tag{32}$$

where endogenous error comes from failing to anticipate the future path $\mathbf{x}_+$ of the solution, and exogenous error comes from failing to anticipate $\mathbf{y}_+$.

Exogenous error can be written as simply $\epsilon_T^x = -[\mathbf{J}^{-1}]_{T,+}\mathbf{y}_+$, where $\mathbf{J}^{-1}$ is the true inverse of $\mathbf{J}$. Endogenous error is somewhat more complex, and can be written as $\epsilon_T^n = [\mathbf{J}^{-1}]_{T,+}\mathbf{J}_{+,T}\widehat{\mathbf{x}}_T$. Here, $\mathbf{J}_{+,T}\widehat{\mathbf{x}}_T$ captures how the truncated $\widehat{\mathbf{x}}_T$ has a persistent effect on future $\mathbf{y}$, and $[\mathbf{J}^{-1}]_{T,+}$ is the endogenous change in $\mathbf{x}_T$ that is necessary to offset this.

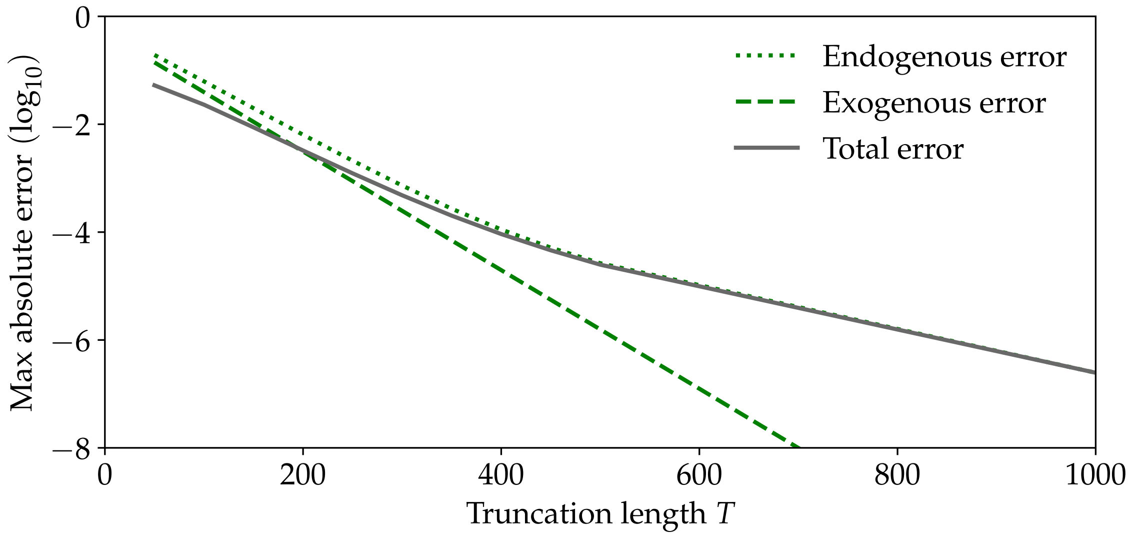


Figure 11: Approximation errors for output response to deficit-financed tax cut

Figure 11 shows this error decomposition as a function of truncation length for the intertemporal Keynesian cross equation $\mathbf{A}d\mathbf{Z} = d\mathbf{B}$, in response to a deficit-financed tax cut. Here, as in the rest of this section, we use an alternative calibration of the model that features more endogenous persistence in order to test the limits of truncation. All details of the calibration and shock are discussed in appendix B.6.

We see that for short T, endogenous and exogenous errors are large and have similar magnitudes, and indeed partly cancel each other. But for longer T, the exogenous error falls to minimal levels as the original fiscal shock recedes, while the endogenous error falls more slowly and eventually drives the entire approximation error. This is common in models with strong internal persistence.

Practical difficulties with truncation. We know from proposition 12 that the truncated solution eventually approaches the correct one. Indeed, truncation usually works well—but as figure 11 reveals, a long truncation horizon T can sometimes be necessary to obtain an accurate enough answer. Solving the system with such a T can be costly in its own right, especially in the block case. It can be even costlier to identify what T is high enough in the first place, since this generally involves solving the system with different T and verifying that changes in T make little difference (see Auclert et al. 2021).

Another issue is that we often must multiply several Jacobians when setting up our system, like in (26), prior to solving it. When performed with truncated matrices, this process introduces additional error from missing anticipation, and although choosing high enough T usually alleviates any problems, there is no analog here of the theoretical guarantee from proposition 12. Indeed, there are examples, like (30), where this process fails catastrophically, and the Jacobian calculated via truncation is not invertible even when the true Jacobian is.

Both these issues motivate us to seek alternatives to simple truncated matrix operations, which we tackle in the next two subsections.

4.2 Iteration using approximate inverses

With a truncated system $\mathbf{J}_{T,T}\mathbf{x}_T = \mathbf{y}_T$, we can directly obtain the solution $\widehat{\mathbf{x}}_T$ using standard methods. The cost of doing this grows cubically in T, making it difficult to use large T.

Alternatively, we can use *iterative methods*. When solving iteratively for $\widehat{\mathbf{x}}_T$, we do not explicitly perform Gaussian elimination on $\mathbf{J}_{T,T}$. Instead, we only need to evaluate $\mathbf{J}_{T,T}$ on sequences $\widehat{\mathbf{x}}_T$. This has a far lower cost that grows quadratically in T, making it practical to choose high enough T that truncation error is avoided. To be able to iterate, however, we need an approximate inverse $\mathbf{P}_{T,T} \approx \mathbf{J}_{T,T}^{-1}$, which in the context of iterative methods is called a *preconditioner*.

Fortunately, for quasi-Toeplitz $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$, there is a natural choice of $\mathbf{P}$: the Toeplitz part $T(\mathbf{j}^{-1})$ of the inverse. This is correct modulo a compact operator, and very cheap to calculate, since the symbol satisfies $j^{-1}(z) = j(z)^{-1}$. (See appendix A.3 for details.)

If desired, we can augment $\mathbf{P}$ as follows. We take the inverse of $\mathbf{J}_{2\tilde{T},2\tilde{T}}$ for some small $\tilde{T}$, such that the inverse is inexpensive to compute, and find the non-Toeplitz part of the first $\tilde{T} \times \tilde{T}$ submatrix of this inverse. We can use this as an approximation to the compact correction of the true inverse, and add it to $\mathbf{P}$.

Given either preconditioner $\mathbf{P}$, we can then iterate over $\widehat{\mathbf{x}}_T$ in two ways:

- *Simple iteration*: starting with initial guess $\widehat{\mathbf{x}}_T^0 = 0$, we iteratively evaluate

$$\widehat{\mathbf{x}}_T^l = \widehat{\mathbf{x}}_T^{l-1} + \mathbf{P}_{T,T}(\widehat{\mathbf{y}}_T - \mathbf{J}_{T,T}\widehat{\mathbf{x}}_T^{l-1}) \tag{33}$$

  to obtain a sequence of guesses $\widehat{\mathbf{x}}_T^l$, until $\widehat{\mathbf{y}}_T - \mathbf{J}_{T,T}\widehat{\mathbf{x}}_T^l$ is sufficiently close to zero. This uses the approximate inverse $\mathbf{P}_{T,T}$ to correct the error from the previous iteration.

- *Generalized minimal residual method (GMRES)*: defining the lth *Krylov subspace* to be the span of $(\mathbf{P}_{T,T}\mathbf{J}_{T,T})^m\mathbf{P}_{T,T}\mathbf{y}_T$ for $m = 0, \ldots, l-1$, the lth guess $\widehat{\mathbf{x}}_T^l$ is chosen from this subspace to minimize the norm of the residual $\mathbf{y}_T - \mathbf{J}_{T,T}\widehat{\mathbf{x}}_T^l$, again continuing until the residual is sufficiently close to zero.

Preconditioner	Iterative strategy	
	Simple	GMRES
Toeplitz part of inverse	10	4
With compact correction ($\tilde{T} = 100$)	6	4
With compact correction ($\tilde{T} = 200$)	4	3

Table 1: Number of iterations until max absolute residual $< 10^{-8}$

GMRES is a standard iterative method used to solve large-scale linear models, and is widely implemented (Saad 2003). It works especially well when the preconditioned matrix $\mathbf{P}_{T,T}\mathbf{J}_{T,T}$ is close to the identity, at least up to some error with rank that is not too high. This is plausible in our case, since the difference between $\mathbf{PJ}$ and $\mathbf{I}$ is a compact operator, and compact operators can be uniformly approximated by finite-rank matrices.[31]

GMRES is always at least as good as simple iteration, since the lth iterate in (33) is contained in the lth Krylov subspace. In practice, GMRES can be vastly more efficient.

Application to fiscal shock. We now apply these iterative methods to solve for impulse responses to the fiscal shock introduced in the previous section. We use both simple iteration and GMRES, and try three preconditioners $\mathbf{P}$: first, the Toeplitz part of the symbol $T(\mathbf{j}^{-1})$ on its own, and then also adding approximate compact corrections with $\tilde{T} = 100$ and $\tilde{T} = 200$. In all these strategies, we solve for a sequence of length $T = 3000$.

Table 1 shows the number of iterations in each case that are necessary to obtain a solution with residual under 10^{-8}. We see convergence is reasonably quick in all cases, taking a maximum of 10 iterations with simple iteration and a pure-Toeplitz preconditioner, and falling to only 4 iterations when the preconditioner is augmented with 200-by-200 compact correction.

The performance of simple iteration is more sensitive to including the compact correction in $\mathbf{P}$: it improves from 10 to 4 iterations, while GMRES takes a maximum of 4 iterations even with the pure-Toeplitz $\mathbf{P}$, and falls to 3 with a corrected $\mathbf{P}$. This is because GMRES can be very effective even when $\mathbf{PJ}$ differs from the identity, so long as the error has low rank, whereas simple iteration requires $\mathbf{PJ}$ to be reasonably close to the identity. Indeed, in more complex models, simple iteration can sometimes fail to converge at all.

[31] Of course, there is also an error in $\mathbf{P}_{T,T}\mathbf{J}_{T,T}$ caused by truncation, but with this method it is cheap to choose T high enough that this error is minimal.

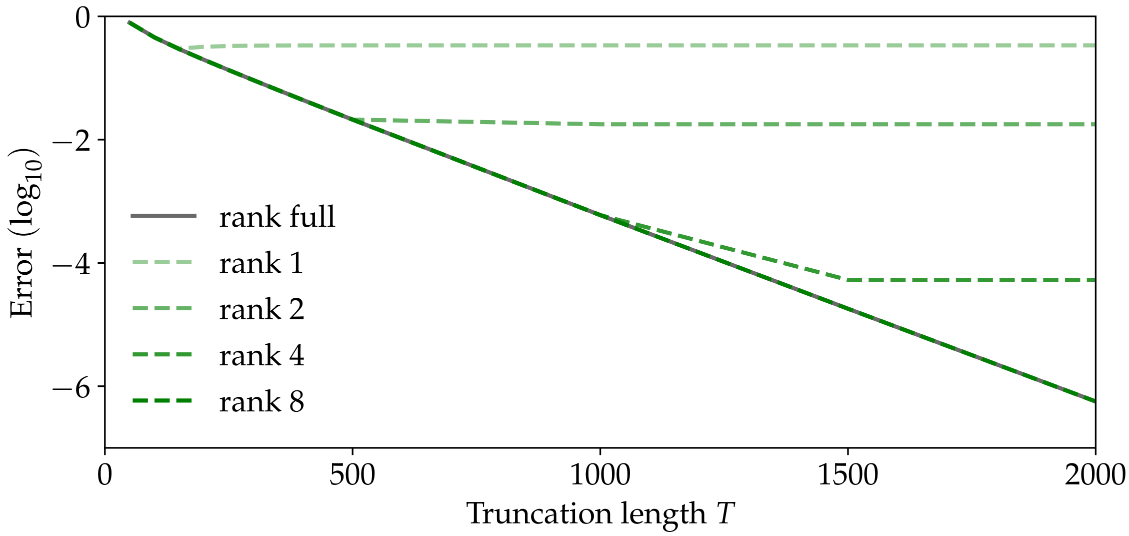


Figure 12: Error in evaluating $\mathbf{A}d\mathbf{Y}$ for compressed correction matrices

4.3 Directly using quasi-Toeplitz structure

In the last section, we exploited quasi-Toeplitz structure as part of an iterative strategy. We now discuss how this structure can be used directly, to avoid the cost of storing and doing computations with large truncated matrices. In this part we rely heavily on the insights of Bini et al. (2019).

There are two key observations. First, it is cheap to work with Toeplitz operators. For instance, if we truncate $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ to a horizon of T, then for the Toeplitz part $T(\mathbf{j})$ we need only store the length $2T-1$ sequence $j_{-(T-1)}, \dots, j_0, \dots, j_{T-1}$. Especially when T is large, this is far smaller than a $T \times T$ matrix. Applying this Toeplitz part to a sequence is also cheap, and can be done in $O(T \log T)$ time with the fast Fourier transform (see appendix A.3), rather than the usual $O(T^2)$ for matrix-vector multiplication.

Of course, this is not useful if we still need to work with a dense $T \times T$ correction matrix $\mathbf{E}_{T,T}$. This brings us to the second key observation. Since $\mathbf{E}$ is compact, it can be arbitrarily well-approximated by finite rank operators—and in practice, this approximation can often be good with surprisingly low rank. Hence, rather than working with a $T \times T$ truncated matrix $\mathbf{E}_{T,T}$, we can work with a low-rank factorization $\mathbf{E}_{T,T} \approx \mathbf{U}_{T,r}\mathbf{V}'_{T,r}$, where $\mathbf{U}_{T,r}$ and $\mathbf{V}_{T,r}$ have only $r \ll T$ columns each. Storing this factorization requires only $2Tr$ elements rather than T^2, and applying it to a sequence has cost $O(Tr)$ rather than $O(T^2)$. Further, since $E_{t,s}$ decays as t, s move away from $(0,0)$, we can potentially choose a lower T.

Figure 12 illustrates this second point for the same heterogeneous-agent model as in the last section. It plots the error in applying the asset Jacobian $\mathbf{A}$ to a geometrically-decaying income shock $d\mathbf{Y} \equiv \{1, 0.9, 0.9^2, \dots\}$, when $\mathbf{A}$ is represented as the sum of a precise Toeplitz part (truncated to $T = 3000$) and an approximate correction. The horizontal axis gives the T to which the correction is truncated, and each line corresponds to approximation with a different rank r.[32]

In this example, we see that a very high T is needed for accuracy. This is because the model has been intentionally chosen to feature very high internal persistence. (See appendix B.6.) Importantly, however, the rank can be *far* lower than the truncation length T. For instance, when the correction is truncated to $T = 2000$, a rank-8 approximation delivers error indistinguishable from using the full $T \times T$ matrix.

Furthermore, this accuracy of low-rank approximations is not an artifact of any particular model, but instead seems to hold across a variety of heterogeneous-agent models. In the appendix, we provide two examples. First, in appendix D.2 we show that in the quantitative version of the Intertemporal Keynesian Cross model from Auclert et al. (2024a), which combines a two-account household block with frictional investment, price and wage rigidities, and capital gains, the general equilibrium mapping from government spending shocks to output is well-approximated with a low-rank compact correction. In an analog of figure 12, we obtain accuracy to 8 digits with a rank-10 correction, and 10 digits with a rank-15 correction. Second, in appendix D.3 we show that in the quantitative menu cost models from Auclert et al. (2024c), both the "pass-through matrix" mapping nominal marginal costs to prices and the "generalized Phillips curve" mapping real marginal costs to inflation are quasi-Toeplitz with very low-rank corrections, with at most five singular values above 10^{-10}.

To summarize, together with the iterative methods in the previous section, we now have a recipe for much more efficient sequence-space operations, which are feasible even when T is high. First, rather than using direct methods to solve a sequence-space system, with cost $O(T^3)$, we can do just a few iterations, each of which requires a matrix-vector product with cost of only $O(T^2)$. Second, by splitting Jacobians into Toeplitz and compact parts, we can reduce the latter cost further to just $O(T(\log T + r))$, for some low rank r.

5 Iterative solutions for large-scale models

To demonstrate the effectiveness of iterative methods in handling large systems, we now consider a many-country generalization of the intertemporal Keynesian cross model we

[32] Here, we use the r leading singular values from a direct singular value decomposition (SVD) to construct the low-rank approximation. To avoid the cost of a full SVD, one can use Lanczos iteration. (See Bini et al. 2019.)

have studied so far, in which countries purchase part of their consumption baskets from each other. Related models are derived in Aggarwal, Auclert, Rognlie and Straub (2023) and Bellifemine, Couturier and Jamilov (2025); here we extend these models to a large, asymmetric trade network.

Description of model. Here, the linearized goods-market clearing condition in country i is

$$d\mathbf{Y}_i = \sum_j d\mathbf{C}_{i,j} + d\mathbf{G}_i, \tag{34}$$

where $d\mathbf{Y}_i$ and $d\mathbf{G}_i$ are the sequences of GDP and government consumption changes in country i, and $d\mathbf{C}_{i,j}$ is the sequence containing the amount of country i's good consumed by country j in each period.

We assume a fixed fraction $\Pi_{i,j}$ of country j's total consumption is imported from country i (i.e. $d\mathbf{C}_{i,j} = \Pi_{i,j} d\mathbf{C}_j$). We also assume that the monetary authority in each country targets a constant real interest rate r. Country j's total consumption $d\mathbf{C}_j$ is then determined by its households' total post-tax income $d\mathbf{Y}_j - d\mathbf{T}_j$ via the domestic intertemporal MPC matrix $\mathbf{M}_j$, which implies the relationship

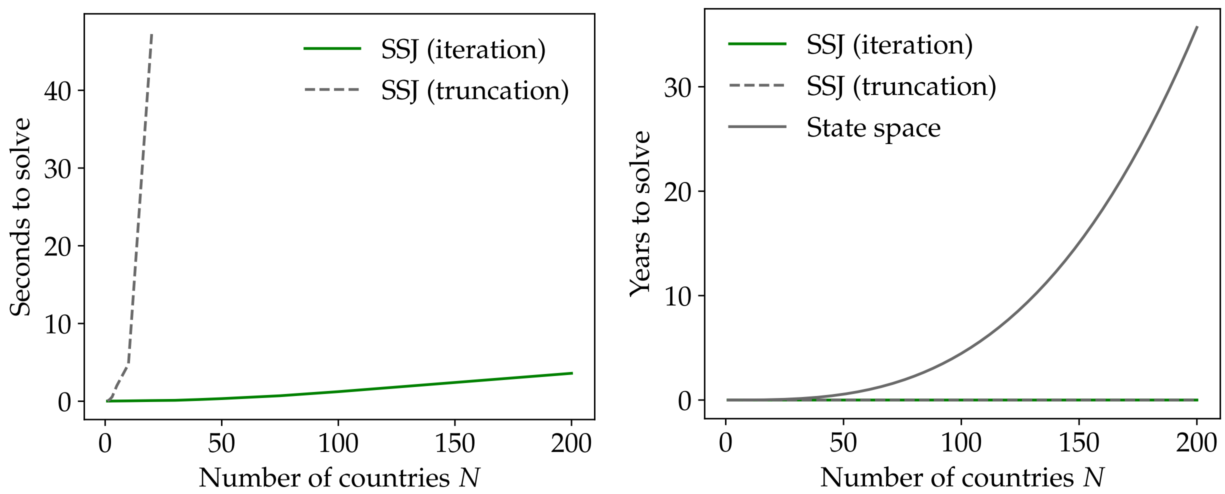
$$d\mathbf{Y}_i = \sum_j \Pi_{i,j} \mathbf{M}_j (d\mathbf{Y}_j - d\mathbf{T}_j) + d\mathbf{G}_i, \tag{35}$$

which is the international generalization of the intertemporal Keynesian cross. We assume that fiscal policy $d\mathbf{T}_i, d\mathbf{G}_i$ in each country is chosen exogenously, consistent with each government's intertemporal budget constraint.

To calibrate a realistic trade network, we use data from the BACI database of bilateral trade flows.[33] For $i \neq j$, we calibrate $\Pi_{i,j}$ to be the value of country j's imports from country i, divided by country j's net purchases, i.e. its GDP less net exports ($\Pi_{i,j} = M_{i,j}/(GDP_j - NX_j)$); we calibrate the diagonal entries as $\Pi_{i,i} = 1 - \sum_{j \neq i} \Pi_{j,i}$, so that $\mathbf{\Pi}$ is column-stochastic. After excluding countries with a ratio of imports to GDP less net exports above 1, the matrix of trade flows yields our calibration of $\mathbf{\Pi}$ for the 190 countries in our sample.[34] Lacking information on intertemporal MPCs for each country, we set all $\mathbf{M}_j$ equal to the $\mathbf{M}$ in our main calibration.

[33] A more detailed description of this dataset is provided in Gaulier and Zignago (2010).

[34] Our calibration excludes countries with a total imports to GDP less net exports ratio above 1 (Cambodia, Marshall Islands, St. Lucia, Seychelles, Vietnam, and Tuvalu) as well as Cuba and North Korea, which are exceedingly isolated. The few very high import ratios in the data are likely due to imported intermediate inputs, which are not in our model.



(a) Iteration v. truncation

(b) Sequence space v. state space

Figure 13: Time taken to solve international IKC model

We then assume the US implements the same deficit-financed tax cut shock as in the previous section, and study its propagation in both the US and the rest of the world.

Computation. Equation (35) implies that output in each country depends on output in every other country, thanks to consumption-income feedback. It is therefore necessary to solve for output in each country simultaneously.[35] With 190 countries and a truncation horizon of $T = 1000$, the sequence-space system we must solve is of size 190000×190000, which is far too large for a direct solution.

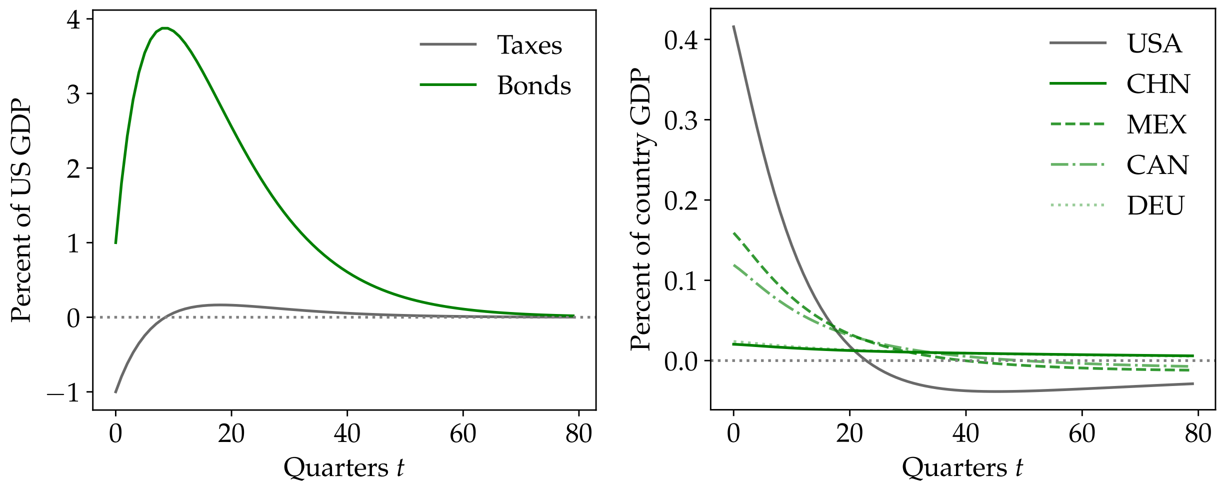
Instead, we apply the methods from the previous section. Concretely, we stack (35) across countries, replace it for one country by the world asset-market clearing condition, and apply GMRES with the Toeplitz part of the inverse as the preconditioner, following the method in section 4.2. This strategy proves to be immensely successful, taking 12 iterations and fewer than 3 seconds on a laptop to converge.

Figure 13(a) compares the performance of this iterative method to the conventional, direct method of solving a truncated linear system, for calibrations with varying numbers of countries.[36] Before we even hit 30 countries, the direct method takes almost a minute, while the iterative method takes less than a second. The gap widens from there.

The gap is even larger when we compare to state-space methods à la Reiter (2009), which solve for a recursive law of motion in the distribution. We depict the estimated

[35]In Aggarwal et al. (2023), it was possible to exploit symmetry to decouple (35) between countries. With our asymmetric calibration of $\mathbf{\Pi}$ taken from the data, we are not aware of any similar simplification.

[36]For this exercise, we randomly generate $\mathbf{\Pi}$s with some home bias.



(a) Response of fiscal variables to shock

(b) Response of GDP

Figure 14: Response to deficit-financed tax shock in the US

time necessary to solve the same system using these methods in the "state space" line of figure 13(b).[37] Models that take years to solve in the state space take seconds in the sequence space. This illustrates the combined power of the sequence space, which reduces the complexity of heterogeneous-agent models to a collection of sequences, and iterative methods for sequence-space solutions.

The global effects of a shock to the US. Figure 14(b) plots the resulting changes in output for a few key countries. The fiscal shock in the US is plotted on the left in figure 14(a).

On impact, all countries experience a boom, but the size of the boom in each country depends on how tightly linked it is with the US. For countries like Canada and Mexico, which are among the US's larger trade partners, there is a larger boom than in countries that trade less with the US. Canada and Mexico, unlike the US's more distant trade partners, also experience a hangover after the initial boom, as economic activity in the US itself slows once taxes are raised to start paying down the debt. Appendix E.2 visualizes output responses around the world both on impact and after 20 quarters.

[37] We estimate the time needed for a state-space solution by taking the size of the state-space system implied by our model with N countries and timing a Schur decomposition for a matrix of that dimension, which captures the key bottleneck step—separating into stable and unstable eigensystems—in the state-space method. For N where it is no longer feasible to perform the Schur decomposition, we extrapolate using its cubic complexity.

6 Conclusion

We have shown that sequence-space Jacobians have a special *quasi-Toeplitz* structure. This structure delivers a simple test for existence and uniqueness of solutions, and also opens the door to more efficient and accurate computation, which allows us to solve sequence-space systems of unprecedented size.

Going forward, a top priority will be further exploration of these new computational methods. It may be possible to adapt them to solve other models with many aggregate variables in the sequence space, such as spatial models of trade or migration, or heterogeneous-household models with an input-output structure as in Schaab and Tan (2023). Another important extension is to better handle models with unit roots, such as the partial equilibrium representative-agent model of consumption, which we currently exclude by working in ℓ^2.

Separately, even with the improvements in section 4.3, there is also still a great deal of redundancy in our representation of Jacobians. A more parsimonious representation, perhaps with rational functions, may make it possible to solve even larger models.

References

Acharya, Sushant and Keshav Dogra, "Understanding HANK: Insights from a PRANK," *Econometrica*, 2020, *88* (3), 1113–1158.

Aggarwal, Rishabh, Adrien Auclert, Matthew Rognlie, and Ludwig Straub, "Excess Savings and Twin Deficits: The Transmission of Fiscal Stimulus in Open Economies," *NBER Macroeconomics Annual*, 2023, *37* (1), 325–412.

Ahn, SeHyoun, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf, "When Inequality Matters for Macro and Macro Matters for Inequality," *NBER Macroeconomics Annual*, 2018, *32* (1), 1–75.

Aiyagari, S Rao, "Uninsured Idiosyncratic Risk and Aggregate Saving," *The Quarterly Journal of Economics*, 1994, *109* (3), 659–684.

Auclert, Adrien and Matthew Rognlie, "Inequality and Aggregate Demand," 2020.

__ , Bence Bardóczy, Matthew Rognlie, and Ludwig Straub, "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," Working Paper 26123, National Bureau of Economic Research 2019.

__ , __ , __ , and __ , "Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models," *Econometrica*, 2021, *89* (5), 2375–2408.

__ , Matthew Rognlie, and Ludwig Straub, "Micro Jumps, Macro Humps: Monetary Policy and Business Cycles in an Estimated HANK Model," Working Paper 26647, National Bureau of Economic Research 2020.

__ , __ , and __ , "The Intertemporal Keynesian Cross," *Journal of Political Economy*, 2024, *132* (12), 4068–4121.

— , — , **and** — , "Fiscal and Monetary Policy with Heterogeneous Agents," *Annual Review of Economics*, 2025, *17.*

— **, Michael Cai, Matthew Rognlie, and Ludwig Straub**, "Optimal Long-Run Fiscal Policy with Heterogeneous Agents," 2024.

— **, Rodolfo Rigato, Matthew Rognlie, and Ludwig Straub**, "New Pricing Models, Same Old Phillips Curves?," *The Quarterly Journal of Economics*, 2024, *139* (1), 121–186.

Axler, Sheldon, *Measure, Integration & Real Analysis*, Springer Nature, 2020.

Baxter, Marianne and Robert G King, "Productive Externalities and Business Cycles," Technical Report, Federal Reserve Bank of Minneapolis 1991.

Bellifemine, Marco, Adrien Couturier, and Rustam Jamilov, "The Regional Keynesian Cross," *Working Paper*, 2025.

Benhabib, Jess and Roger EA Farmer, "Indeterminacy and Increasing Returns," *Journal of Economic Theory*, 1994, *63* (1), 19–41.

Bilbiie, Florin O, "Monetary Policy and Heterogeneity: An Analytical Framework," *Review of Economic Studies*, 2024, p. rdae066.

Bini, Dario A, Stefano Massei, and Leonardo Robol, "Quasi-Toeplitz Matrix Arithmetic: A MATLAB Toolbox," *Numerical Algorithms*, 2019, *81*, 741–769.

Blanchard, Olivier Jean and Charles M Kahn, "The Solution of Linear Difference Models Under Rational Expectations," *Econometrica*, 1980, pp. 1305–1311.

Boppart, Timo, Per Krusell, and Kurt Mitman, "Exploiting MIT Shocks in Heterogeneous-Agent Economies: The Impulse Response As a Numerical Derivative," *Journal of Economic Dynamics and Control*, 2018, *89*, 68–92.

Böttcher, Albrecht and Bernd Silbermann, *Introduction to Large Truncated Toeplitz Matrices*, Springer Science & Business Media, 2012.

— **and Sergei M Grudsky**, *Spectral Properties of Banded Toeplitz Matrices*, SIAM, 2005.

Carroll, Christopher D, Edmund Crawley, Jiri Slacalek, Kiichi Tokuoka, and Matthew N White, "Sticky Expectations and Consumption Dynamics," *American Economic Journal: Macroeconomics*, 2020, *12* (3), 40–76.

Clancey, Kevin F and Israel Gohberg, *Factorization of Matrix Functions and Singular Integral Operators*, Vol. 3, Birkhäuser, 1981.

Gabaix, Xavier, "A Behavioral New Keynesian Model," *American Economic Review*, 2020, *110* (8), 2271–2327.

Gaulier, Guillaume and Soledad Zignago, "BACI: International Trade Database at the Product-Level. The 1994-2007 Version," Working Papers 2010-23, CEPII 2010.

Gilli, Manfred and Giorgio Pauletto, "Krylov Methods for Solving Models with Forward-Looking Variables," *Journal of Economic Dynamics and Control*, 1998, *22* (8-9), 1275–1289.

Hagedorn, Marcus, "Local Determinacy in Incomplete-Markets Models," Technical Report, CEPR Discussion Papers 2023.

Hansen, Lars Peter and Thomas J Sargent, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 1980, *2*, 7–46.

__ **and** __ , "Linear Rational Expectations Models for Dynamically Interrelated Variables," Technical Report, Federal Reserve Bank of Minneapolis 1980.

Hazell, Jonathon, Juan Herreño, Emi Nakamura, and Jón Steinsson, "The Slope of the Phillips Curve: Evidence from US States," *The Quarterly Journal of Economics*, 2022, *137* (3), 1299–1344.

Huo, Zhen and Naoki Takayama, "Rational Expectations Models with Higher-Order Beliefs," *Review of Economic Studies*, 2024, p. rdae096.

Hurwicz, Leonid and Mark Walker, "On the Generic Nonoptimality of Dominant-Strategy Allocation Mechanisms: A General Theorem That Includes Pure Exchange Economies," *Econometrica*, 1990, *58* (3), 683–704.

Krusell, Per and Anthony A Smith Jr, "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Luenberger, David G, *Optimization by Vector-Space Methods*, John Wiley & Sons, 1997.

Meyer-Gohde, Alexander, "Solving and Analyzing DSGE Models in the Frequency Domain," Technical Report, IMFS Working Paper Series 2024.

Mrkaic, Mico, "Policy Iteration Accelerated with Krylov Methods," *Journal of Economic Dynamics and Control*, 2002, *26* (4), 517–545.

Onatski, Alexei, "Winding Number Criterion for Existence and Uniqueness of Equilibrium in Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2006, *30* (2), 323–345.

Ravn, Morten O and Vincent Sterk, "Macroeconomic Fluctuations with HANK & SAM: An Analytical Approach," *Journal of the European Economic Association*, 2021, *19* (2), 1162–1202.

Reiter, Michael, "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, 2009, *33* (3), 649–665.

__ , "Approximate and Almost-Exact Aggregation in Dynamic Stochastic Heterogeneous-Agent Models," 2010.

Saad, Yousef, *Iterative Methods for Sparse Linear Systems*, SIAM, 2003.

Schaab, Andreas and Stacy Yingqi Tan, "Monetary and Fiscal Policy According to HANK-IO," *Unpublished Manuscript*, 2023.

Schmitt-Grohé, Stephanie and Martın Uribe, "Closing Small Open Economy Models," *Journal of international Economics*, 2003, *61* (1), 163–185.

Sims, Christopher A, "Solving Linear Rational Expectations Models," *Computational Economics*, 2002, *20* (1-2), 1.

__ , "On the Genericity of the Winding Number Criterion for Linear Rational Expectations Models," 2007.

Tan, Fei and Todd B Walker, "Solving Generalized Multivariate Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 2015, *60*, 95–111.

Uhlig, Harald, "A Toolkit for Analysing Nonlinear Stochastic Models Easily," 1995.

Weinstein, Jonathan and Muhamet Yildiz, "A Structure Theorem for Rationalizability with Application to Robust Predictions of Refinements," *Econometrica*, 2007, *75* (2), 365–400.

Whiteman, Charles H, "Spectral Utility, Wiener-Hopf Techniques, and Rational Expectations," *Journal of Economic Dynamics and Control*, 1985, *9* (2), 225–240.

Wolf, Christian K., "Interest Rate Cuts versus Stimulus Payments: An Equivalence Result," *Journal of Political Economy*, 2025, *133* (4), 1235–1275.

A Practical computation

In this appendix we discuss practical computation of the winding number and symbol. For quasi-Toeplitz computational questions not covered here, a useful reference is Bini et al. (2019).

It is also worth noting that computing a full Jacobian is unnecessary to obtain $j(z) = j_{-(T-1)}z^{-(T-1)} + \ldots + j_0 + \ldots + j_{T-1}z^{T-1}$. One can also numerically evaluate, out to date $2T-1$, the impulse response to a small MIT shock that shocks the input at the single date $T-1$ (for sufficiently high T), and store the impulse response as the coefficients $j_{-(T-1)}, \ldots, j_{T-1}$. Careful numerical differentiation (e.g. two-sided), and potentially some extrapolation as in the second part of appendix A.2, can be helpful to make this computation well-behaved.

A.1 Winding number

To numerically calculate the winding number $\text{wind}(j)$ of the symbol $j(z)$ associated with some quasi-Toeplitz Jacobian $\mathbf{J}$, we follow two basic steps:

- **Step 1.** Evaluate $j(z)$ at all roots of unity $z_k \equiv e^{2\pi ik/K}$, for some large even K (as a rule of thumb, usually 5000–10,000 or more).
- **Step 2.** Test how many times the piecewise linear path formed by $j(z_0), j(z_1), \ldots, j(z_{K-1}), j(z_0)$ winds counterclockwise around the origin.

In the block case where the symbol $j(z)$ is matrix-valued, we take the determinant $\det j(z_k)$ at each z_k at the end of step 1, then calculate how many times the piecewise path formed by $\det j(z_k)$ winds in step 2.

Below we discuss details of implementing both steps.[A-1] Code is available at `https://github.com/shade-econ/nber-workshop-2025/blob/main/notebooks/winding_number.py`. Before continuing, it is worth making a few additional points. First, a full Jacobian is not necessary to implement this test. In principle, we can get the coefficients of $j(z)$ by numerically finding the response to a sufficiently well-anticipated shock. Second, to obtain the symbol for a heterogeneous-agent block from the Jacobian, see appendix A.2.

Step 1 details. A quasi-Toeplitz $\mathbf{J}$ may be specified in many ways. If it is defined by multiplying, inverting, or linearly combining other quasi-Toeplitz Jacobians, it is best to

[A-1] These are fairly standard computations, not original to this paper; we include them here for reference.

perform step 1 individually for these individual Jacobians, then performing the corresponding operations on the results. For instance, if $\mathbf{J} = (\mathbf{J}_1\mathbf{J}_2 - 4\mathbf{J}_3)^{-1}\mathbf{J}_4$, then $j(z) = (j_1(z)j_2(z) - 4j_3(z))^{-1}j_4(z)$, and we can evaluate j at all z_k by evaluating j_1, j_2, j_3, j_4 and then calculating $(j_1(z_k)j_2(z_k) - 4j_3(z_k))^{-1}j_4(z)$. (This holds identically in the block case, with matrix rather than scalar multiplication and inversion.) See section 3.6 for an example of this.

For a Jacobian that cannot be further broken down in this way, we proceed as follows. We assume that we have some truncated representation $j(z) = j_{-(T-1)}z^{-(T-1)} + \ldots + j_0 + \ldots + j_{T-1}z^{T-1}$, and also assume that K is even and $K \geq 2T$. We observe that $\widehat{j}(z) \equiv z^{K/2}j(z)$ is then a polynomial of degree less than K and can be written in the form $\widehat{j}(z) = \widehat{j}_0 + \widehat{j}_1 z + \ldots + \widehat{j}_{K-1}z^{K-1}$. Applying the fast Fourier transform (FFT) to the sequence of K coefficients $\widehat{j}_0, \ldots, \widehat{j}_{K-1}$ gives the sequence $\widehat{j}(z_0), \widehat{j}(z_{K-1}), \ldots, \widehat{j}(z_1)$ from evaluating the polynomial $\widehat{j}(z)$ at clockwise Kth roots of unity. Since $z_k^{K/2} = (-1)^k$, we have $j(z_k) = (-1)^k\widehat{j}(z_k)$, and can obtain $j(z_0), j(z_1), \ldots, j(z_{K-1}), j(z_0)$ by multiplying $\widehat{j}(z_0), \widehat{j}(z_{K-1}), \ldots, \widehat{j}(z_1)$ by the sequence $1, -1, 1, -1, \ldots$, adding $\widehat{j}(z_0)$ at the end, and reversing the sequence.

To summarize, the algorithm is:

- **Inputs:** sequence $j_{-(T-1)}, \ldots, j_{T-1}$ for symbol; even $K \geq 2T$
- Initialize sequence $\widehat{j}_0, \ldots, \widehat{j}_{K-1}$ to zero
- Set $\widehat{j}_{k+K/2} \equiv j_k$ for $k = -(T-1), \ldots, T-1$
- Take FFT of sequence $\widehat{j}_0, \ldots, \widehat{j}_{K-1}$, store output as $x_0, \ldots, x_{K-1}$
- **Return:** sequence $x_0, -x_{K-1}, x_{K-2}, \ldots, -x_1, x_0$, which gives values of symbol j on counterclockwise Kth roots of unity.

For the block case where the input sequence $j_{-(T-1)}, \ldots, j_{T-1}$ are $n \times n$ matrices, we do the above for each pair of indices $n_1, n_2 \in \{1, \ldots, n\}$, returning a sequence of $n \times n$ matrices, and then taking the determinant of each matrix to reduce to a scalar sequence.

Because the FFT is efficient, with cost $O(T \log T)$, the above algorithm is generally quite fast. It is possible to further economize on the FFT and any other operations (including the determinant, or e.g. multiplication if we are composing Jacobians) by taking advantage of conjugate symmetry: since $z_k = z_{K-k}^*$, where $*$ denotes the complex conjugate, we have $j(z_{K-k}) = j(z_k)^*$ and $\det j(z_{K-k}) = (\det j(z_k))^*$ as well, making it unnecessary to compute these separately.

Step 2 details. Once we have our sample values $j(z_0), j(z_1), \ldots, j(z_{K-1}), j(z_0)$ (or their determinants in the block case), which we will denote by $x_0 + iy_0, \ldots, x_K + iy_K$, we need to see how many times the closed piecewise linear path formed by the coordinates (x_k, y_k) winds counterclockwise around $(0, 0)$. It is often useful to assess this visually.

There are many ways, however, to automate this evaluation. One is to count how many times the path crosses the ray $(0, 0) \to (\infty, 0)$ from below (counting crossings from above negatively). We summarize an implementation of this as follows:

- **Input:** Closed path $(x_0, y_0), (x_1, y_1), \ldots, (x_{K-1}, y_{K-1}), (x_K, y_K)$
- Initialize $w \equiv 0$, $s_0 \equiv 1_{y_0 \geq 0}$
- For $k = 1, \ldots, K$:
  - Calculate $s_k \equiv 1_{y_k \geq 0}$. If $s_k = s_{k-1}$, then there is not a crossing. If $s_k \neq s_{k-1}$, evaluate the following three cases:
    - If $x_k, x_{k-1} > 0$, then there is a crossing. Add $s_k - s_{k-1}$ to w.
    - If $x_k, x_{k-1} \leq 0$, then there is not a crossing.
    - If $x_k > 0$ and $x_{k-1} \leq 0$ or vice versa, then there may or maybe not be a crossing. If $(x_{k-1}y_k - x_k y_{k-1})/(y_k - y_{k-1}) > 0$, then segment connecting (x_{k-1}, y_{k-1}) and (x_k, y_k) crosses x-axis to the right of origin and there is a crossing: add $s_k - s_{k-1}$ to w. Otherwise, there is not.
- **Return:** winding number w

Assuming that the loop is done in compiled code, this algorithm is extremely fast—especially since in most practical examples, we only have $s_k \neq s_{k-1}$ a few times in the loop $k = 1, \ldots, K$.

At https://github.com/shade-econ/nber-workshop-2025/blob/main/notebooks/winding_number.py, the `sample_values_simple()` function implements an easy-to-understand version of step 1 above for the scalar case, while `sample_values()` accommodates the block case as well and takes advantage of conjugate symmetry for efficiency. `winding_number_of_path()` implements step 2.

Further improvements are possible: for instance, if the symbol decays exponentially past some point, with $j_t = C\rho^{t-\tau}$ for some C and $t \geq \tau$, then we can analytically calculate the contribution of $j_t z^t + j_{t+1} z^{t+1} + \ldots$ to the symbol, which is simply $\frac{Cz^t}{1-\rho z}$. This is only useful if we want to extrapolate beyond $t = K/2$, though, which is fairly rare if we are sampling a large number K of z_k in step 1 to begin with.

A.2 Calculating symbol for heterogeneous-agent blocks

A key input to the algorithm for step 1 in the previous section is the truncated sequence $j_{-(T-1)}, \ldots, j_{T-1}$ describing the symbol, for Jacobians that cannot be further reduced to simpler ones. This is trivial for simple macroeconomic equations, but more subtle for the heterogeneous-agent blocks from section 2.5, which we cover here.

From proposition 3, we know that j_k is the sum of all entries (out to infinity) on the kth lower diagonal in the fake news matrix $\boldsymbol{\mathcal{F}}$. If we compute a $T \times T$ Jacobian from a $T \times T$ fake news matrix, following the algorithm from Auclert et al. (2021), then the final diagonal entries on the bottom and right of the matrix give sums of all available entries in $\boldsymbol{\mathcal{F}}$ on each respective diagonal, which can approximate the infinite sum. This gives us a simple way to extract a symbol.

- **Simple approach:** given a $T \times T$ truncated Jacobian calculated for a heterogeneous-agent block from the fake news algorithm, extract the symbol as

$$j_k \equiv \begin{cases} J_{T-1+k,T-1} & -(T-1) \leq k \leq 0 \\ J_{T-1,T-1-k} & 0 \leq k \leq T-1. \end{cases} \tag{A.1}$$

This approach is *not* a good idea for arbitrary truncated Jacobians, because they may have truncation artifacts ("spurious missing anticipation") near T as discussed in section 4.1. But het-agent Jacobians as calculated in Auclert et al. (2021) are built up from the fake news matrix and do not have these artifacts.

The simple approach can work well, but the downside is that each j_k only sums finitely many entries of $\boldsymbol{\mathcal{F}}$, and near the ends this sum is especially short: for instance, j_{T-1} is just the single term $J_{T-1,0} = \mathcal{F}_{T-1,0}$. If the block's internal persistence is large enough relative to T, then the resulting symbol can be a poor approximation to the truth.

An alternative approach, starting with the same data, is to use "double extrapolation". First, away from the matrix's edges, we exponentially extrapolate entries of the fake news matrix along the diagonal to obtain a better approximation to the infinite sum. Then, we exponentially extrapolate the symbol on the left and right. Below we describe this procedure in more detail:

- **Double-extrapolation approach:** given a $T \times T$ Jacobian and the associated fake news matrix, and some positive τ, calculate

$$f_k \equiv \begin{cases} \mathcal{F}_{T-1+k,T-1} & -(T-1-\tau) \leq k \leq 0 \\ \mathcal{F}_{T-1,T-1-k} & 0 \leq k \leq T-1-\tau. \end{cases} \tag{A.2}$$

and

$$\rho_k \equiv \begin{cases} \mathcal{F}_{T-1+k,T-1}/\mathcal{F}_{T-2+k,T-2} & -(T-1-\tau) \leq k \leq 0 \\ \mathcal{F}_{T-1,T-1-k}/\mathcal{F}_{T-2,T-2-k} & 0 \leq k \leq T-1-\tau. \end{cases} \tag{A.3}$$

i.e. where f_k is the last entry of the fake news matrix on lower diagonal k, and ρ_k is the rate of growth of that entry vs. the one before it on the same diagonal. Then add j_k as defined in (A.1) to the extrapolated correction $f_k \frac{\rho_k}{1-\rho_k}$ to obtain our improved coefficients $\bar{j}_k$.

We now have $\bar{j}_k$ for k from $-\bar{T}$ through $\bar{T}$, where $\bar{T} \equiv T-1-\tau$. We can then exponentially extrapolate out as far as desired, by defining $\bar{j}_k \equiv \bar{j}_{\bar{T}} \cdot (\bar{j}_{\bar{T}}/\bar{j}_{\bar{T}-1})^{k-\bar{T}}$ for $k > \bar{T}$ on the right and similarly $\bar{j}_k \equiv \bar{j}_{-\bar{T}} \cdot (\bar{j}_{-\bar{T}}/\bar{j}_{-(\bar{T}-1)})^{-\bar{T}-k}$ for $k < -\bar{T}$ on the left.

The double-extrapolation approach takes advantage of the fact that at long horizons, anticipation and persistence become close to exponential, reflecting the dominant root of the underlying system. We have found that it works well for models with high internal persistence, like the calibration we use in section 4, and avoids the downsides of brute truncation. A choice of $\tau \gg 0$, so we do not extrapolate from $\mathcal{F}$ entries too close to 0 (which do not necessarily decay exponentially), is important. We generally choose τ of at least 200 or 300.

For robustness, one can also define ρ_k in (A.3) using the decay over a longer span—the last 5, 10, or 50 entries on the diagonal, rather than the immediately preceding entry—although in our own applications we have found that (A.3) works fine. It is important, however, to begin extrapolation before the entries of $\boldsymbol{\mathcal{F}}$ are so small that they are affected by numerical error, so the T itself should not be too large. Here, one can check that the log decay of $\boldsymbol{\mathcal{F}}$ along a few diagonals is relatively steady, and does not become either erratic or flatten out by T (both of which indicate likely numerical issues).

A.3 Obtaining inverse symbol and Toeplitz-vector multiplication

Obtaining inverse symbol. In section 4.2, we advocate an iterative strategy that uses the Toeplitz part of the inverse, $T(\mathbf{j}^{-1})$, as a preconditioner.

To obtain this inverse, we proceed as follows, building on what we already know how to calculate. Start with the information produced from step 1 of appendix A.1: $j(z)$ evaluated on counterclockwise Kth roots of unity z_k, for some large even K. Then invert to obtain $j(z_k)^{-1} = j^{-1}(z_k)$. Flip the order so that the roots are clockwise, multiply each by $(-1)^k$, then apply the inverse FFT, obtaining some output $x_0, \ldots, x_{K-1}$. Finally, obtain coefficients of the inverse symbol $(j^{-1})_k \equiv x_{k+K/2}$ for any desired $k = -(T-1), \ldots, T-1$, where $2T < K$.[A-2]

Toeplitz-vector multiplication. In sections 4.2 and 4.3, we also take advantage of the fact that Toeplitz-vector multiplication is cheap. This is due to a standard procedure called "circulant embedding".

The calculation can be briefly described as follows. Suppose that we have a vector $\mathbf{x} = x_0, \dots, x_{T-1}$, and we are multiplying it by $T(\mathbf{j})$ with truncated $\mathbf{j} = j_{-(T-1)}, \dots, j_0, \dots, j_{T-1}$. We take the FFT of the sequence $j_0, \dots, j_{T-1}, j_{-(T-1)}, \dots, j_{-1}$, and then the FFT of $x_0, \dots, x_{T-1}$ padded with zeros at the end to make it also length $2T-1$. Finally, we multiply these two FFTs and take the inverse FFT of the result to get $y_0, \dots, y_{2T-1}$. The first T entries give our answer $\mathbf{y} = y_0, \dots, y_{T-1} = T(\mathbf{j})\mathbf{x}$.

B Appendix to section 2

B.1 Proof of lemma 1 and proposition 1

These are both standard results on Toeplitz operators.

Proof of lemma 1. We note that we can write any Toeplitz $T(\mathbf{j})$ as a series in the forward operator $\mathbf{F}$ and lag operator $\mathbf{L}$:

$$T(\mathbf{j}) = \sum_{k=1}^{\infty} j_{-k}\mathbf{F}^k + \sum_{k=0}^{\infty} j_k \mathbf{L}^k$$

Since $\mathbf{F}$ and $\mathbf{L}$, and their powers, are all bounded operators with norms of 1, we can write

$$\|T(\mathbf{j})\| \leq \sum_{k=1}^{\infty} |j_{-k}| \|\mathbf{F}^k\| + \sum_{k=0}^{\infty} |j_k| \|\mathbf{L}^k\| = \sum_{k=-\infty}^{\infty} |j_k| \equiv \|\mathbf{j}\|_1 \tag{A.4}$$

and thus $T(\mathbf{j})$ is bounded, with finite norm of at most $\|\mathbf{j}\|_1$.

Proof of proposition 1. This result is usually stated in terms of Hankel operators (see, e.g., Böttcher and Grudsky 2005).

To see it for ourselves without the Hankel machinery, we note that in (9), we already showed that $C_{s+j,s}$ approaches c_j as $s \to \infty$, where the sequence $\mathbf{c} = \{c_j\}$ is defined to be the convolution of $\mathbf{a}$ and $\mathbf{b}$. The convolution theorem for z-transforms tells us that

[A-2] Like before, this is unchanged in the block case, if we take $j(z_k)^{-1}$ to be a matrix inverse, and perform the inverse FFT in parallel for each $i, j = 1, \dots, n$.

$c(z) = a(z)b(z)$. Further, we showed in (11) that the difference $\mathbf{E} = \mathbf{C} - T(\mathbf{c})$ is given by $E_{t,s} = -\sum_{k=1}^{\infty} a_{t+k} b_{-s-k}$. All that remains is to show that this $\mathbf{E}$ is compact.

To do this, we define $\mathbf{E}_{\tilde{a},\tilde{b}} \equiv T(\tilde{\mathbf{a}})T(\tilde{\mathbf{b}}) - T(\tilde{\mathbf{a}} * \tilde{\mathbf{b}})$, for arbitrary absolutely summable two-sided sequences $\tilde{\mathbf{a}}, \tilde{\mathbf{b}}$. We claim that the map from $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ (with 1-norms) to $\mathbf{E}_{\tilde{a},\tilde{b}}$ (with operator norm) is bilinear and bounded. Since bilinearity is mechanical, we will focus on boundedness. First, from (A.4) above, we already know that the maps from $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ to $T(\tilde{\mathbf{a}})$ and $T(\tilde{\mathbf{b}})$ are bounded. Second, one can show that the mapping from $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ to the convolution $\tilde{\mathbf{a}} * \tilde{\mathbf{b}}$ is bounded:

$$\|\tilde{\mathbf{a}} * \tilde{\mathbf{b}}\|_1 = \sum_n \left| \sum_k \tilde{a}_k \tilde{b}_{n-k} \right| \leq \sum_k |\tilde{a}_k| \sum_n |\tilde{b}_{n-k}| = \|\tilde{\mathbf{a}}\|_1 \|\tilde{\mathbf{b}}\|_1. \tag{A.5}$$

We combine this with (A.4) to get a bounded bilinear mapping from $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ to $T(\tilde{\mathbf{a}} * \tilde{\mathbf{b}})$, and thus to $\mathbf{E}_{\tilde{a},\tilde{b}}$.

Now, for any n, define $\mathbf{a}^n$ and $\mathbf{b}^n$ to be "truncated" versions of $\mathbf{a}$ and $\mathbf{b}$, with all entries with absolute index greater than n set to zero. It follows from the formula in (11) that all entries of $\mathbf{E}_{a^n,b^n}$ with row or column index greater than n are zero. Since $\mathbf{E}_{a^n,b^n}$ has finitely many nonzero entries, it has finite rank.

Clearly, $\mathbf{a}^n$ and $\mathbf{b}^n$ approach $\mathbf{a}$ and $\mathbf{b}$ as $n \to \infty$. Since the mapping to $\mathbf{E}_{a^n,b^n}$ is bounded and therefore continuous, it follows that $\mathbf{E}_{a^n,b^n} \to \mathbf{E}_{a,b} \equiv \mathbf{E}$ as $n \to \infty$. We can therefore approximate $\mathbf{E}$ arbitrarily well (in sup norm) with finite-rank operators, implying that it is compact.

B.2 Proof of proposition 3

First, we note that for $t, s > 0$, we have:

$$\mathcal{F}_{t,s} \equiv J_{t,s} - J_{t-1,s-1} = E_{t,s} - E_{t-1,s-1}. \tag{A.6}$$

We can therefore write

$$\begin{aligned}
\sum_{k=1}^{\infty} \mathcal{F}_{t+k,s+k} &= \sum_{k=1}^{\infty} (E_{t+k,s+k} - E_{t+k-1,s+k-1}) \\
&= -E_{t,s} + \lim_{k\to\infty} E_{t+k,s+k} = -E_{t,s}
\end{aligned}$$

where the second line follows from the telescoping sum, and $\lim_{k\to\infty} E_{t+k,s+k} = 0$ from the compactness of $\mathbf{E}$. This proves the right side of (15). We now observe that for $k \geq 0$,

$$j_k = J_{k,0} - E_{k,0} = \mathcal{F}_{k,0} + \sum_{v=1}^{\infty} \mathcal{F}_{k+v,v} = \sum_{v=0}^{\infty} \mathcal{F}_{k+v,v},$$

proving the left side of (15) (together with a symmetric argument for $k < 0$).

Finally, suppose that we have an $\boldsymbol{\mathcal{F}}$ whose entries have a finite absolute sum. Then for $t \geq s$, we recursively sum to obtain $J_{t,s} = \sum_{k=0}^{s} \mathcal{F}_{t-s+k,k}$. If we define $\mathbf{j}$ and $\mathbf{E}$ following (15), we note that

$$J_{t,s} = \sum_{k=0}^{\infty} \mathcal{F}_{t-s+k,k} - \sum_{k=s+1} \mathcal{F}_{t-s+k,k} = j_s + E_{t,s}$$

verifying (together with a symmetric argument for $t < s$) that $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$.

B.3 Fake news matrix for a product of Toeplitz matrices

We found in (10) that when taking the product $\mathbf{C} = T(\mathbf{a})T(\mathbf{b})$, we have $\mathcal{F}_{t,s} \equiv C_{t,s} - C_{t-1,s-1} = a_t b_{-s}$ for all $t, s > 0$. All that remains is to characterize the case where $t = 0$ or $s = 0$. There, by definition, we have $\mathcal{F}_{t,s} = C_{t,s}$, which by (8) equals $\sum_{u=0}^{\infty} a_{t-u} b_{u-s}$. Combining these, we have

$$\mathcal{F}_{t,s} = \begin{cases} a_t b_{-s} & t, s > 0 \\ \sum_{u=0}^{\infty} a_{t-u} b_{u-s} & t = 0 \text{ or } s = 0 \end{cases} \tag{A.7}$$

We can interpret (A.7) as follows. For $t, s > 0$, $\mathcal{F}_{t,s}$ is the pure effect at date t of having anticipated at date 0 that there would be a shock at date s. This is the anticipatory effect b_{-s}, propagated forward by a_t.

If $s = 0$, then $\mathcal{F}_{t,0}$ is the total effect at t from a surprise shock at date 0, which is $\sum_{u=0}^{\infty} a_{t-u} b_u$. This sums up all lagged effects b_u, propagated either forward or backward to t by a_{t-u}.

If $t = 0$, then $\mathcal{F}_{0,s}$ is the total effect at date 0 from anticipating a shock at s, which is $\sum_{u=0}^{\infty} a_{-u} b_{u-s}$. This sums up anticipation a_{-u} at every horizon of possible lead or lag effects b_{u-s} of the shock.

Special case: lower-triangular $T(\mathbf{a})$, upper-triangular $T(\mathbf{b})$. If $T(\mathbf{a})$ is lower-triangular (purely backward-looking, with $a_k = 0$ for $k < 0$) and $T(\mathbf{b})$ is upper-triangular (purely forward-looking, with $a_k = 0$ for $k > 0$), then the only nonzero term in the sum $\sum_{u=0}^{\infty} a_{t-u} b_{u-s}$

is $a_t b_{-s}$, corresponding to $u = 0$. (A.7) then simplifies to just

$$\mathcal{F}_{t,s} = a_t b_{-s} \tag{A.8}$$

for all t, s. In other words, the fake news matrix is just the outer product of the nonzero parts of $\mathbf{a}$ and $\mathbf{b}$. This is true, for instance, in our Calvo pricing example, where $a_t = (1-\theta)\theta^t$ for $t \geq 0$ and 0 for $t < 0$, and $b_{-s} = (1-\beta\theta)(\beta\theta)^s$ for $s \geq 0$ and 0 for $s < 0$.

This case occurs fairly often, since we often have a purely forward-looking Toeplitz matrix $T(\mathbf{b})$ that governs policy (e.g. reset pricing), and then multiply this by a purely backward-looking Toeplitz matrix $T(\mathbf{a})$ that governs the evolution of the state (e.g. the aggregate price). For instance, for the bond-in-utility (BU) model of consumption and savings, Auclert et al. (2024a) derive an asset Jacobian that is the product of a backward-looking Toeplitz matrix with first column $(1, \lambda, \lambda^2, \ldots)$ and a forward-looking Toeplitz matrix with first row $(1-m, -m(\beta\lambda), -m(\beta\lambda)^2, \ldots)$. The fake news matrix for this asset Jacobian is just the outer product of $(1, \lambda, \lambda^2, \ldots)$ and $(1-m, -m(\beta\lambda), -m(\beta\lambda)^2, \ldots)$.

We can even regard the formula in (A.13) for the $t \geq 1$ heterogeneous-agent fake news matrix as a block version of this case, with $b_{-s} \equiv \frac{\partial \mathbf{D}_1}{\partial X_s}$ and $a_t \equiv \mathcal{E}_{t-1}$.

B.4 Deviations from rational expectations and quasi-Toeplitz structure

Auclert et al. (2020) shows how Jacobians derived under rational expectations can be manipulated to obtain the Jacobians implied by various behavioral or information frictions. One simple way to embed a variety of frictions in a common framework is to define $\Theta_{t,s}$ to be the fraction of a change relative to steady state at time s that is anticipated at time t; for dX_s, for instance, the expectation at time t would $\Theta_{t,s} dX_s$. (We generally assume $\Theta_{t,s} = 1$ for all $t \geq s$, i.e. that all current or past dX_s are fully known.) For instance, for $t > s$, cognitive discounting as in Gabaix (2020) corresponds to $\Theta_{t,s} = \alpha^{t-s}$, and sticky expectations as in Carroll, Crawley, Slacalek, Tokuoka and White (2020) corresponds to $\Theta_{t,s} = 1 - \theta^{t+1}$. We refer to $\mathbf{\Theta} \equiv (\Theta_{t,s})$ as the *expectations matrix*.

The approach of Auclert et al. (2020) can be summarized in this setting by saying that the modified Jacobian under non-rational expectations, $\widetilde{\mathbf{J}}$, has entries

$$\widetilde{J}_{t,s} = \sum_{\tau=0}^{\min(t,s)} \Theta_{\tau,s} \mathcal{F}_{t-\tau,s-\tau}, \tag{A.9}$$

where $\boldsymbol{\mathcal{F}}$ is the fake news matrix under rational expectations.[A-3] Intuitively, $\mathcal{F}_{t-\tau,s-\tau}$ gives

[A-3]See, for instance, the slides at `https://github.com/shade-econ/nber-workshop-2023/blob/main/`

the effect at date t from having anticipated at date τ that there would be a change at date s, and in (A.9) we sum these times the amount $\Theta_{\tau,s}$ of anticipation at each τ.

Applying definition (14), the fake news matrix $\widetilde{\boldsymbol{\mathcal{F}}}$ corresponding to the modified Jacobian can be written as

$$\widetilde{\mathcal{F}}_{t,s} = \widetilde{J}_{t,s} - \widetilde{J}_{t-1,s-1} = \Theta_{0,s}\mathcal{F}_{t,s} + \sum_{\tau=1}^{\min(t,s)} (\Theta_{\tau,s} - \Theta_{\tau-1,s-1})\mathcal{F}_{t-\tau,s-\tau}. \tag{A.10}$$

Now, under weak restrictions on $\boldsymbol{\Theta}$, we can extend the absolute summability of $\boldsymbol{\mathcal{F}}$—which sections 2.5 and 2.6 prove is true for broad classes of models under rational expectations—to $\widetilde{\boldsymbol{\mathcal{F}}}$ as well.

Lemma 3. *Suppose that $\boldsymbol{\mathcal{F}}$ is absolutely summable, the entries of $\boldsymbol{\Theta}$ are bounded, and that there is also some bound M such that*

$$\sum_{\tau=1}^{\infty} |\Theta_{\tau,s+\tau} - \Theta_{\tau-1,s+\tau-1}| < M \tag{A.11}$$

for all s, i.e. that the sum of absolute first differences is uniformly bounded across all diagonals. Then $\widetilde{\boldsymbol{\mathcal{F}}}$ is also absolutely summable, and therefore by proposition 3 $\widetilde{\mathbf{J}}$ is quasi-Toeplitz.

Proof. We can rearrange the absolute sum over all t and s of $\widetilde{\mathcal{F}}_{t,s}$ in (A.10) to be

$$\sum_{t=0}^{\infty}\sum_{s=0}^{\infty} |\Theta_{0,s}||\mathcal{F}_{t,s}| + \sum_{t'=0}^{\infty}\sum_{s'=0}^{\infty} |\mathcal{F}_{t',s'}| \cdot \sum_{\tau=1}^{\infty} |\Theta_{\tau,s'+\tau} - \Theta_{\tau-1,s'+\tau-1}|, \tag{A.12}$$

where we derive the second term by summing the second term in (A.10) over all t and s, substituting $t' = t - \tau$ and $s' = s - \tau$, and rearranging.

The absolute summability of the first term in (A.12) follows from the absolute summability of $\boldsymbol{\mathcal{F}}$ and the boundedness of $\boldsymbol{\Theta}$, and the absolute summability of the second term follows from the absolute summability of $\boldsymbol{\mathcal{F}}$ and the uniform boundedness assumption in (A.11). ∎

The two conditions on $\boldsymbol{\Theta}$, boundedness and (A.11), are satisfied in every practical case we have encountered. (A.11) is necessary only to rule out pathological cases, e.g. where $\Theta_{t,s} = 1$ for all odd s and $\Theta_{t,s} = 0$ for all even s, which would break the result.

Lectures/Lecture%2010%20Information%20handout.pdf or lecture notes at https://mrognlie.github.io/econ411-3/econ411_3_lecture10.pdf for more discussion of (A.9), and also some more detail on the mapping of frictions to $\boldsymbol{\Theta}$. In these notes we use "$\mathbf{E}$" (already taken in this paper for the compact correction) rather than "$\boldsymbol{\Theta}$".

(A.11) is also easily implied by simpler sufficient conditions, such as assuming that the entries of $\boldsymbol{\Theta}$ are weakly increasing as we go down diagonals.[A-4] (This corresponds to assuming that agents have weakly more information about what will happen a given number of periods in the future, if it has been longer since period 0 when the shock was realized.) We summarize this as the lemma below.

Lemma 4. *Suppose that $\boldsymbol{\mathcal{F}}$ is absolutely summable, the entries of $\boldsymbol{\Theta}$ are bounded, and the entries of $\boldsymbol{\Theta}$ are weakly increasing along diagonals. Then $\boldsymbol{\mathcal{F}}$ is absolutely summable and $\widetilde{\mathbf{J}}$ is quasi-Toeplitz.*

Proof. If entries are weakly increasing along diagonals and bounded, then

$$\sum_{\tau=1}^{\infty} |\Theta_{\tau,s+\tau} - \Theta_{\tau-1,s+\tau-1}| = \sum_{t=1}^{\infty} \Theta_{\tau,s+\tau} - \Theta_{\tau-1,s+\tau-1} = \lim_{\tau\to\infty} \Theta_{\tau,s+\tau} - \Theta_{0,s},$$

which is uniformly bounded by the boundedness assumption on $\boldsymbol{\Theta}$, implying (A.11) and thus the result. ∎

Finally, although our result applies to very general $\boldsymbol{\Theta}$, it is worth noting that not all deviations from full information and rational expectations can be summarized by a $\boldsymbol{\Theta}$: for instance, with learning, expectations about $dX_{s'}$ may depend on the realized value at dX_s. Although we are optimistic that our structure results will extend to these cases as well, we leave this to future work.

B.5 Proof of proposition 4

So that the Jacobian and fake news matrix are well-defined, we will assume local differentiability of all mappings (16)–(18) around the steady state. Auclert, Bardóczy, Rognlie and Straub (2021) show that the t, s entry of the fake news matrix can be written as

$$\mathcal{F}_{t,s} = \begin{cases} \frac{\partial Y_0}{\partial X_s} & t = 0 \\ \mathcal{E}'_{t-1} \frac{\partial \mathbf{D}_1}{\partial X_s} & t \geq 1 \end{cases} \tag{A.13}$$

where $\frac{\partial Y_0}{\partial X_s}$ and $\frac{\partial \mathbf{D}_1}{\partial X_s}$ are the responses of the aggregate outcome at date 0 and incoming distribution at date 1 to the shock at date s (written as dY_0^s and $d\mathbf{D}_1^s$ in Auclert et al. 2021), and $\mathcal{E}_{t-1} \equiv (\boldsymbol{\Lambda}_{ss})^{t-1}\mathbf{y}_{ss}$ is the $t-1$th expectation function.

[A-4] The same argument goes through if $\boldsymbol{\Theta}$ is weakly decreasing along diagonals, or if it is weakly monotonic for some $t \geq T$ and $s \geq S$.

We note that for $s \geq 1$, we can write

$$\frac{\partial Y_0}{\partial X_s} = (y_v (v_v)^{s-1} v_X)' \mathbf{D}_{ss} \tag{A.14}$$
$$\frac{\partial \mathbf{D}_1}{\partial X_s} = (\Lambda_v (v_v)^{s-1})' \mathbf{D}_{ss} \tag{A.15}$$

where v_v and v_X are the derivatives of the function v in (16) with respect to v and X around the steady state, y_v is the derivative of y in (18) with respect to v, and so on.

By stationarity, v_v has all eigenvalues strictly inside the unit circle, implying that for some $\Delta_1 < 1$ that bounds these eigenvalues and some constant C_1, we have $\|(v_v)^s\| \leq C_1 \Delta_1^s$ for all s. It follows from (A.14)–(A.15) that we can choose suitable constants C_2 and C_3 such that

$$\left\| \frac{\partial Y_0}{\partial X_s} \right\| \leq C_2 \Delta_1^s \quad \text{and} \quad \left\| \frac{\partial \mathbf{D}_1}{\partial X_s} \right\| \leq C_3 \Delta_1^s \tag{A.16}$$

for all s. (We can always choose C_2 and C_3 such that this bound applies for $s = 0$ as well.)

By stationarity, $\mathbf{\Lambda}_{ss}$ has all eigenvalues except one inside the unit circle. Since $\mathbf{\Lambda}_{ss}$ is row-stochastic, the other eigenvalue must be 1, corresponding to $\mathbf{\Lambda}_{ss}\mathbf{1} = \mathbf{1}$. If we define $\mathcal{C}\mathbf{x} \equiv \mathbf{x} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{x}$ to be the *demeaning* operator that returns a vector minus its mean, then we note that $\mathcal{C}\mathbf{\Lambda}_{ss}$ has the same eigenvalues as $\mathbf{\Lambda}_{ss}$, but with the 1 replaced by a 0, and thus its eigenvalues are all strictly inside the unit circle and bounded by some $\Delta_2 < 1$. Furthermore, for any t, we have $(\mathcal{C}\mathbf{\Lambda}_{ss})^t = \mathcal{C}(\mathbf{\Lambda}_{ss})^t$. We conclude that we can choose some C_4 such that $\|\mathcal{C}\mathcal{E}_t\| = \|(\mathcal{C}\mathbf{\Lambda}_{ss})^t \mathbf{y}_{ss}\| \leq C_4 \Delta_2^t$.

Finally, we observe that since the sum of the distribution never changes, $\mathbf{1}'\frac{\partial \mathbf{D}_1}{\partial X_s} = 0$, and therefore $(\mathcal{C}\mathcal{E}_{t-1})'\frac{\partial \mathbf{D}_1}{\partial X_s} = \mathcal{E}_{t-1}'\mathcal{C}'\frac{\partial \mathbf{D}_1}{\partial X_s} = \mathcal{E}_{t-1}'\frac{\partial \mathbf{D}_1}{\partial X_s}$, allowing us to bound with

$$\left\| \mathcal{E}_{t-1}' \frac{\partial \mathbf{D}_1}{\partial X_s} \right\| \leq C_4 \Delta_2^{t-1} C_3 \Delta_1^s \equiv C_5 \Delta_2^t \Delta_1^s. \tag{A.17}$$

Finally, going back to (A.13), and combining (A.16) and (A.17), if we set $C \equiv \max(C_2, C_5)$ we obtain

$$|\mathcal{F}_{t,s}| \leq C \Delta_2^t \Delta_1^s$$

which, setting $\Delta \equiv \max(\Delta_2, \Delta_1)$, is the desired bound.

B.6 Description of heterogeneous-agent model and fiscal shock

Model and calibration. We take a standard heterogeneous-agent household block, where individual households make consumption and savings decisions subject to uninsurable idiosyncratic risk and an occasionally binding borrowing constraint, as in Auclert et al.

(2024a) and Auclert et al. (2025). Most of our calibration choices are standard: we have a quarterly calibration with an elasticity of intertemporal substitution of 1, an AR(1) lognormal income process with annualized persistence 0.91 and cross-sectional standard deviation 0.92 (both from Auclert et al. 2024a), and an annualized real interest rate of 2%. As in the baseline model in Auclert et al. (2024a), we assume that the only asset is government debt.

The remaining parameter to calibrate is the discount factor β. We do this in two ways, both of which hit a quarterly weighted MPC target of 0.2 as in Auclert et al. (2025):

- **Calibration 1.** This is a simple, standard calibration, which we use as our calibration outside of section 4. Here we have a single β for all agents, and calibrate it to $\beta = 0.9734$ to hit the quarterly MPC target. This implies a fairly low aggregate asset level, equal to 68% of annual household income.

- **Calibration 2.** Here, loosely inspired by Auclert et al. (2025), we also attempt to target a more realistic level of aggregate assets, equal to 500% of annual household income. Here, we do so by replacing the single β with a stochastic β, where as in Auclert et al. (2025), households draw a new, random β on average every 100 quarters (25 years) to reflect generational turnover. We allow for four equispaced β values, and calibrate the level and spacing of the βs to hit both the MPC and asset targets. The resulting calibration features a minimum $\beta = 0.9445$ and a maximum $\beta = 1.0023$. (Note that since this maximum β is not permanent, it is permissible despite being greater than 1.) This more sophisticated calibration implies a richer state space—since households are heterogeneous in income, assets, and patience—and also features much more endogenous persistence, since the most patient households try to hold on to their assets indefinitely. Due to these more challenging features, we use it as our calibration in section 4, so that we have a testbed where approximating the symbol and correction is relatively difficult.

Countercyclical income inequality. For our determinacy exercises in section 3, we also allow for an alternative version of the model where income inequality (and thus also income risk) is countercyclical in aggregate income. Concretely, as in Auclert and Rognlie (2020) and Auclert et al. (2025), we assume that aggregate labor demand N_t is rationed across households with an "incidence function" $n_{it} = \gamma(N_t, e_{it}) = N_t \frac{e_{it}^{\zeta \log N_t}}{\mathbb{E} e_{it}^{\zeta \log N_t}}$, where $\mathbb{E}$ denotes cross-sectional expectations. Households' total pretax income is then $n_{it}e_{it}$, where e_{it} is their exogenous productivity. Since e_{it} becomes negatively correlated with n_{it} when aggregate $N_t = Y_t$ is high, inequality is countercyclical, with ζ being the sensitivity of the standard deviation of log income with respect to aggregate output.

In this case, the Jacobians $\mathbf{A}$ and $\mathbf{M}$ out of income changes from $Y_t = N_t$ changing are in principle different from the Jacobians $\mathbf{A}^T$ and $\mathbf{M}^T$ out of income changes from taxes T_t changing. Since we only consider the $\zeta \neq 0$ case when analyzing determinacy around the steady state—holding taxes and government spending unchanged—this distinction does not matter for us here.

Fiscal shock. For our fiscal shock of interest in sections 4 and 5, we choose a deficit-financed tax cut similar to Auclert et al. (2025). We specify the tax-cut shock itself, $dT_t^{shock} \equiv -0.9^t$, to decay geometrically. We then write $dB_t = 0.975 \cdot dB_{t-1} - dT_t^{shock}$, so that all else equal, debt in a period increases by the amount of the tax-cut shock, but that the fiscal rule also attempts to reduce the total debt by 2.5% each quarter. The actual tax path is then determined residually from $\{dB_t\}$: $dT_t = (1+r)dB_{t-1} - dB_t$. We assume that government spending is zero and unchanged.

B.7 Proof of proposition 5

Step 1: Obtaining the fake news matrix and Jacobian from the state-space solution. As appendix B.8 shows, the state-space solution obtained from (21) is

$$x_t = \mathbf{P}x_{t-1} + \sum_{j=0}^{\infty} \mathbf{R}^j \mathbf{Q}_0 \mathbb{E}_t[u_{t+j}]. \tag{A.18}$$

Then, from it, we can obtain the block entries of the fake news matrix and Jacobian, respectively.

$$\mathcal{F}_{t,s} = \mathbf{P}^t \mathbf{Q}_s = \mathbf{P}^t \mathbf{R}^s \mathbf{Q}_0$$
$$J_{t,s} = \sum_{k=0}^{\min\{t,s\}} \mathbf{P}^{t-k} \mathbf{R}^{s-k} \mathbf{Q}_0$$

For $s = t + j$ and $j \geq 0$ we have that the $(t, t+j)$ entry of the Jacobian is

$$J_{t,t+j} = \left[\sum_{k=0}^{t} \mathbf{P}^k \mathbf{R}^k\right] \mathbf{R}^j \mathbf{Q}_0.$$

For $t = s + j$ and $j \geq 0$ we have that the $(t + j, t)$ entry of the Jacobian is

$$J_{s+j,s} = \mathbf{P}^j \left[\sum_{k=0}^{s} \mathbf{P}^k \mathbf{R}^k \right] \mathbf{Q}_0.$$

Step 2: Deriving the symbol and correction. From (15) in proposition 3, the symbol defining the Toeplitz part is

$$j(z) = \left[\sum_{j=-\infty}^{\infty} z^j \mathbf{P}^{\max\{-j,0\}} \Omega \mathbf{R}^{\max\{j,0\}} \right] \mathbf{Q}_0$$

and the correction is

$$\mathbf{E}_{t,s} = -\mathbf{P}^{\min\{s,t\}+\max\{t-s,0\}+1} \Omega \mathbf{R}^{\min\{s,t\}+\max\{s-t,0\}+1} \mathbf{Q}_0,$$

where Ω is

$$\Omega = \sum_{k=0}^{\infty} \mathbf{P}^k \mathbf{R}^k$$

and satisfies the Lyapunov equation

$$\Omega = \mathbf{P} \Omega \mathbf{R} + \mathbf{I}.$$

This final equation allows us to rewrite the correction as

$$\mathbf{E}_{t,s} = \mathbf{P}^{\min\{s,t\}+\max\{t-s,0\}} (\mathbf{I} - \Omega) \mathbf{R}^{\min\{s,t\}+\max\{s-t,0\}} \mathbf{Q}_0.$$

That $T(\mathbf{j})$ is block Toeplitz and Ω exists follows from the assumption that the solution is stable.

Step 3: Observing the rank of the correction E. We define $s = t + j$. For any possible j, $\min\{t + j, t\} + \max\{-j, 0\} = t$ and $\min\{t + j, t\} + \max\{j, 0\} = t + j$, so

$$\mathbf{E}_{t,s} = \mathbf{P}^t (\mathbf{I} - \Omega) \mathbf{R}^s \mathbf{Q}_0.$$

Define $\overline{\mathbf{P}}$ as all of the $\mathbf{P}^t$ s stacked on top of one another and $\overline{\mathbf{R}}$ as all of the $\mathbf{R}^s$ s stacked beside each other, so that

$$\overline{\mathbf{P}} = \begin{bmatrix} \mathbf{P}^0 \\ \mathbf{P}^1 \\ \mathbf{P}^2 \\ \vdots \end{bmatrix}$$
$$\overline{\mathbf{R}} = \begin{bmatrix} \mathbf{R}^0\mathbf{Q}_0 & \mathbf{R}^1\mathbf{Q}_0 & \mathbf{R}^2\mathbf{Q}_0 & \dots \end{bmatrix}.$$

Now, we observe

$$\mathbf{E} = \overline{\mathbf{P}}\left(\mathbf{I} - \Omega\right)\overline{\mathbf{R}}$$

Clearly, the rank of $\overline{\mathbf{P}}$ is finite ($\mathbf{E}$ is therefore a compact operator on $\ell^2(\mathbb{R}^n)$, meaning $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ is block quasi-Toeplitz) and limited to the rank of $\mathbf{P}$, which is simply the size of the state space.

Finally, one can see, by applying selection matrices to $T(\mathbf{j})$ and $\mathbf{E}$, that the properties of the full Jacobian $\mathbf{J}$ will also apply to the Jacobians that map each exogneous variable to each endogenous variable. That is, they will be quasi-Toeplitz and their corrections will have rank less than or equal to the rank of $\mathbf{P}$.

B.8 Determinacy in the canonical form of (21)

Finding the state-space solution. Start from the following canonical linear system, which is equivalent to that in (21).

$$\mathbf{A}\mathbb{E}_t\left[x_{t+1}\right] + \mathbf{B}x_t + \mathbf{C}x_{t-1} + \mathbf{D}u_t = \mathbf{0}_{n\times 1}.$$

All matrices ($\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$) are $n \times n$ and all vectors x_ts and u_ts are $n \times 1$. If there are more entries in u_t than shocks, we stipulate that all of the residual entries (those not corresponding to shocks in the model) are known to be zero for all t. The solution is

$$x_t = \mathbf{P}x_{t-1} + \sum_{j=0}^{\infty}\mathbf{Q}_j\mathbb{E}_t\left[u_{t+j}\right],$$

with the coefficient matrices satisfying the equations below.

$$\mathbf{0} = \mathbf{A}\mathbf{P}^2 + \mathbf{B}\mathbf{P} + \mathbf{C} \tag{A.19}$$

$$\mathbf{Q}_0 = -\left(\mathbf{A}\mathbf{P} + \mathbf{B}\right)^{-1} \mathbf{D} \tag{A.20}$$

$$\mathbf{Q}_j = \underbrace{\left(-\left(\mathbf{A}\mathbf{P} + \mathbf{B}\right)^{-1} \mathbf{A}\right)^j}_{\mathbf{R}} \mathbf{Q}_0. \tag{A.21}$$

We solve for the coefficients by solving the matrix quadratic (A.19). The process of doing so begins by noting that it can be written in the form

$$\underbrace{\begin{bmatrix} \mathbf{I}_n & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{A} \end{bmatrix}}_{\Phi_1} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{P} \end{bmatrix} \mathbf{P} = \underbrace{\begin{bmatrix} \mathbf{0}_{n\times n} & \mathbf{I}_n \\ -\mathbf{C} & -\mathbf{B} \end{bmatrix}}_{\Phi_0} \begin{bmatrix} \mathbf{I}_n \\ \mathbf{P} \end{bmatrix}.$$

Define the QZ decomposition associated with Γ_0 and Γ_1 such that

$$\Phi_0 = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{0}_{n\times n} & \mathbf{S}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix}^*$$

$$\Phi_1 = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{0}_{n\times n} & \mathbf{T}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{Z}_{11} & \mathbf{Z}_{12} \\ \mathbf{Z}_{21} & \mathbf{Z}_{22} \end{bmatrix}^*,$$

where each block is $n \times n$, $\mathbf{S}_{ii}$ and $\mathbf{T}_{ii}$ are upper triangular, and we've chosen the QZ decomposition that sorts the n smallest generalized eigenvalues into the top right corner. Given this, it's possible to show that $\mathbf{P} = \mathbf{Q}_{11}\mathbf{S}_{11}\mathbf{T}_{11}^{-1}\mathbf{Q}_{11}^{-1} = \mathbf{Z}_{21}\mathbf{Z}_{11}^{-1}$ satisfies (A.19). The $\mathbf{Q}_j$s can then be computed with (A.20) and (A.21).

Under the assumption that this system satisfies the Blanchard-Kahn conditions, this solution exists and is unique. But, what are the Blanchard-Kahn conditions, in this case? We'll cover that now.

Getting to Blanchard and Kahn (1980). First, note that the canonical form in (21) is the same as

$$\Phi_1 \begin{bmatrix} x_t \\ \mathbb{E}_t[x_{t+1}] \end{bmatrix} = \Phi_0 \begin{bmatrix} x_{t-1} \\ x_t \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n\times n} \\ -\mathbf{D} \end{bmatrix} u_t.$$

If we knew $\mathbf{A}$ to be invertible, we could invert Φ_1 and the resulting system would exactly be the canonical form of their equation (1a) in Blanchard and Kahn (1980), where A in

(1a) is equal to $\Phi_1^{-1}\Phi_0$ here. In that case, by their proposition 1, the equivalent condition would be that $\Phi_1^{-1}\Phi_0$ has *exactly* n unstable eigenvalues (i.e. outside the unit circle). Why would that be? We need n extra conditions to pin down the value of $\mathbb{E}_t[x_{t+1}]$, given x_{t-1} and u_t, since we don't know what x_t is. We may impose that x_t is stable, but that only has bite if it could possibly be *unstable*. Consequently, requiring stability only gives us n extra conditions if there are n unstable eigenvalues.

If $\mathbf{A}$ is not known to be invertible, we may use generalized eigenvalues instead to get at the equivalent conditions. A generalized eigenvalue λ of our system solves

$$\det\left(\Phi_0 - \lambda\Phi_1\right) = 0.$$

Then, the conditions for existence and uniqueness are:

1. A rational expectations equilibrium *exists* if there are *at least* n generalized eigenvalues less than 1 in absolute value;

2. A rational expectations equilibrium is *unique* if there are *exactly* n generalized eigenvalues less than 1 in absolute value.

More concretely, the determinant is

$$\det(\Phi_0 - \lambda\Phi_1) = \det\begin{bmatrix} -\lambda\mathbf{I}_n & \mathbf{I}_n \\ -\mathbf{C} & -\mathbf{B} - \lambda\mathbf{A} \end{bmatrix} = \det(\mathbf{A}\lambda^2 + \mathbf{B}\lambda + \mathbf{C}) \tag{A.22}$$

and the generalized eigenvalue λ solves

$$0 = \det(\mathbf{A}\lambda^2 + \mathbf{B}\lambda + \mathbf{C}). \tag{A.23}$$

From Blanchard-Kahn to the winding number. Comparing this to the winding number criterion for (21) we observe a very close resemblance. The determinant of the block symbol for the equivalent block Toeplitz system is

$$\det j(z) = \det(\mathbf{A}z^{-1} + \mathbf{B} + \mathbf{C}z) = z^n \det(\mathbf{A}z^{-2} + \mathbf{B}z^{-1} + \mathbf{C}).$$

Then we recognize that $\operatorname{wind}(\det j)$ is equal to zero only as long as $\det(\mathbf{A}z^{-2} + \mathbf{B}z^{-1} + \mathbf{C})$ has n roots inside the unit circle, since it has $2n$ poles inside the unit circle and z^n has n roots inside the unit circle.

According to the winding number test, we *only* have a unique solution as long as this condition holds. If it has more, a solution does not exist. If it has fewer, there are many solutions. These are precisely the Blanchard-Kahn conditions for the system we described above.

Getting to Sims (2002). Equation (21), in the form of Sims (2002) is

$$\underbrace{\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0}_{n\times n} & \mathbf{I}_n \end{bmatrix}}_{\Gamma_0} \begin{bmatrix} \mathbb{E}_t[x_{t+1}] \\ x_t \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{0}_{n\times n} & -\mathbf{C} \\ \mathbf{I}_n & \mathbf{0}_{n\times n} \end{bmatrix}}_{\Gamma_1} \begin{bmatrix} \mathbb{E}_{t-1}[x_t] \\ x_{t-1} \end{bmatrix} + \underbrace{\begin{bmatrix} -\mathbf{D} \\ \mathbf{0}_{n\times n} \end{bmatrix}}_{\Psi} u_t + \underbrace{\begin{bmatrix} \mathbf{0}_{n\times n} \\ \mathbf{I}_n \end{bmatrix}}_{\Pi} \eta_t, \tag{A.24}$$

where $\eta_t = x_t - \mathbb{E}_{t-1}[x_t]$. In this form, we may solve the model and apply Sims' tests for existence and uniqueness. These are not exactly the same as root-counting tests.

B.9 Details on the neoclassical growth model

Households' problem. Households choose capital K_t and C_t in each period, taking prices as given, to maximize

$$\sum_{t=0}^{\infty} \beta^t U(C_t), \tag{A.25}$$

subject to

$$K_t + C_t \leq (1 + r_t)K_{t-1}. \tag{A.26}$$

For now we assume that $U(C_t) = \frac{C_t^{1-\sigma}}{1-\sigma}$.

Firms take interest rates as given, choosing capital to maximize profits, which are given by

$$\Pi_t = f(K_{t-1}, Z_t) - (1 + r_t)K_{t-1},$$

where $f(K_{t-1}, Z_t) = Z_t K_{t-1}^{\alpha} + (1-\delta)K_{t-1}$.

Combining the households' and firms' FOCs yields the Euler equation

$$0 = U'(f(K_{t-1}, Z_t) - K_t) - \beta f_K(K_t, Z_{t+1})U'(f(K_t, Z_{t+1}) - K_{t+1}), \tag{A.27}$$

which we linearize to obtain (22).

Adding in production externalities. When we add in a production externality à la Baxter and King (1991) and Benhabib and Farmer (1994), the household's problem remains the same, but the firm's problem changes. Firms now take total output and interest rates as given, choosing capital to maximize profits, which are given by

$$\Pi_t = g(Y_t, K_{t-1}, Z_t) - (1+r_t)K_{t-1},$$

where $g(Y_t, K_{t-1}, Z_t) = Z_t Y_t^{\gamma} K_{t-1}^{\alpha} + (1-\delta)K_{t-1}$, with $Y_t = Z_t^{\frac{1}{1-\gamma}} K_t^{\frac{\alpha}{1-\gamma}}$ in equilibrium. Combining these, we define $f(K_{t-1}, Z_t) = Z_t^{\frac{1}{1-\gamma}} K_{t-1}^{\frac{\alpha}{1-\gamma}} + (1-\delta)K_t$, which is $g(\cdot)$ evaluated at the equilibrium level of output.

Now, the modified Euler equation of the neoclassical growth model is

$$0 = U'(f(K_{t-1}, Z_t) - K_t) - \beta\left((1-\gamma)f_K(K_t, Z_{t+1}) + \gamma(1-\delta)\right)U'(f(K_t, Z_{t+1}) - K_{t+1}), \tag{A.28}$$

and the linearized form becomes

$$\begin{aligned} &f_K dK_{t-1} - \left(1 + f_K + (1-\gamma\beta(1-\delta))\frac{f_{KK}}{f_K}\frac{U'}{U''}\right)dK_t + \mathbb{E}_t[dK_{t+1}] \\ &= \left(f_Z + (1-\gamma\beta(1-\delta))\frac{f_{KZ}}{f_K}\frac{U'}{U''}\right)\mathbb{E}_t[dZ_{t+1}] - f_Z dZ_t. \end{aligned} \tag{A.29}$$

Notice that this nests the case without increasing returns.

Why is the rank of the correction to $\mathbf{J}_{dK,dZ}$ still 1?. Start by noting the symbol of $\mathbf{J}_{dK,u} = T(\mathbf{k}) + \mathbf{E}_k$ is $k(z) = \frac{1}{q-p}\sum p^{\max\{k,0\}} q^{\max\{-k,0\}} z^k$. Further, the composite Jacobian is

$$\mathbf{J}_{dK,dZ} = T(\mathbf{k} * \mathbf{h}) + \mathbf{E}_{k,h} + \frac{p}{p-q}\boldsymbol{\mathcal{F}}T(\mathbf{h}), \tag{A.30}$$

where $\mathbf{E}_{k,h} = T(\mathbf{k})T(\mathbf{h}) - T(\mathbf{k} * \mathbf{h})$. Using (11), entry (t,s) of $\mathbf{E}_{k,h}$ is $-\frac{p}{p-q}p^t h_{-1}\mathbf{1}(s<1)$. Notice that this is a rank 1 matrix and that one of the vectors it is composed of is also a basis vector for $\frac{p}{p-q}\boldsymbol{\mathcal{F}}T(\mathbf{h})$. So, we may write the total correction as a rank 1 matrix as well.

C Appendix to section 3

C.1 Proof of proposition 6

This is a standard result. We start with theorem 1.15 from Böttcher and Silbermann (2012) and its proof, which shows that if its winding number is n, we can factorize a Toeplitz operator as

$$T(\mathbf{j}) = \begin{cases} T(\mathbf{j}_-)\mathbf{L}^n T(\mathbf{j}_+) & n \geq 0 \\ T(\mathbf{j}_-)\mathbf{F}^{-n} T(\mathbf{j}_+) & n \leq 0 \end{cases} \tag{A.31}$$

where $T(\mathbf{j}_-)$ and $T(\mathbf{j}_+)$ are invertible operators, and are respectively upper and lower triangular.[A-5]

Due to this invertibility, $\mathrm{indet}(T(\mathbf{j}))$ and $\mathrm{nonex}(T(\mathbf{j}))$ are $\mathrm{indet}(\mathbf{L}^n) = 0$ and $\mathrm{nonex}(\mathbf{L}^n) = n$ for $n \geq 0$, and $\mathrm{indet}(\mathbf{F}^{-n}) = -n$ and $\mathrm{nonex}(\mathbf{F}^{-n}) = 0$ for $n \leq 0$. Cases 2 and 3 of proposition 6 follow.

If $n = 0$, then $\mathrm{indet}(T(\mathbf{j})) = 0$ and $\mathrm{nonex}(T(\mathbf{j})) = 0$, and $T(\mathbf{j})$ is invertible. Let $\mathbf{j}^{-1}$ be the inverse of $\mathbf{j}$ in the sense of convolution.[A-6] We calculate:

$$T(\mathbf{j})^{-1} - T(\mathbf{j}^{-1}) = T(\mathbf{j})^{-1}(\mathbf{I} - T(\mathbf{j})T(\mathbf{j}^{-1})) = -T(\mathbf{j})^{-1}\mathbf{E}$$

where $\mathbf{E}$ is the compact correction from $T(\mathbf{j})T(\mathbf{j}^{-1}) = \mathbf{I} + \mathbf{E}$. Hence $T(\mathbf{j})^{-1}$ equals the Toeplitz $T(\mathbf{j}^{-1})$ plus some compact operator $-T(\mathbf{j})^{-1}\mathbf{E}$, and is quasi-Toeplitz.[A-7]

C.2 Proof of proposition 7

Proposition 6 implies that $\mathbf{J} = T(\mathbf{j})$ is a *Fredholm* operator, since it is an operator on Hilbert space with finite-dimensional kernel and cokernel. It has index

$$\mathrm{ind}(\mathbf{J}) = \mathrm{indet}(\mathbf{J}) - \mathrm{nonex}(\mathbf{J}) = -\mathrm{wind}(j)$$

The proposition then follows from the fact that Fredholm index is unchanged by adding a compact operator $\mathbf{E}$.

C.3 Proof of proposition 8

If proposition 6 applies for some quasi-Toeplitz $\mathbf{J}$, then $\mathbf{J}$ must be injective or surjective. Conversely, if a quasi-Toeplitz $\mathbf{J}$ is injective or surjective, then proposition 7 implies proposition 6 (aside from the structure of $\mathbf{J}^{-1}$ when it exists, which we will handle separately).

[A-5] The corresponding factorization of the symbol into $j_-(z)$, z^n, and $j_+(z)$ is often called the *Wiener-Hopf factorization*.

[A-6] The existence of such an inverse is guaranteed by Wiener's theorem, which states that that for any symbol $j(z) = \sum_{k=-\infty}^{\infty} j_k z^k$ in the Wiener algebra, i.e. where the j_k have a finite absolute sum, then as long as $j(z)$ has no zeros on $\mathbf{T}$, j^{-1} is also in the Wiener algebra. See Example 1.5 of Böttcher and Silbermann (2012).

[A-7] An alternative derivation that explicitly constructs quasi-Toeplitz $T(\mathbf{j})^{-1}$ uses the factorization $T(\mathbf{j}_-)T(\mathbf{j}_+)$ and the fact that the upper and lower triangular parts $T(\mathbf{j}_-)$ and $T(\mathbf{j}_+)$ also have upper and lower triangular Toeplitz inverses, respectively.

Hence, to show that proposition 6 holds for generic $\mathbf{J}$, it suffices to show that $\mathbf{J}$ is generically injective or surjective.

We will make use of the fact that by proposition 7, all quasi-Toeplitz $\mathbf{J}$ where $\text{wind}(j)$ is defined are Fredholm operators on ℓ^2 and thus have closed range.

First, we consider openness. If some $\mathbf{J}$ is injective, then it is a bijection between ℓ^2 and range $\mathbf{J}$, and there is some bounded inverse $\mathbf{K}_0$ defined on range $\mathbf{J}$ such that $\mathbf{K}_0\mathbf{J} = \mathbf{I}$. Because range $\mathbf{J}$ is closed, we can define $\mathbf{K} \equiv \mathbf{K}_0\mathbf{P}$, where $\mathbf{P}$ is the orthogonal projection of ℓ^2 onto range $\mathbf{J}$, to obtain a left inverse $\mathbf{K} : \ell^2 \to \ell^2$ satisfying $\mathbf{KJ} = \mathbf{I}$. Defining $m \equiv \|\mathbf{K}\|^{-1}$, we note that $\|\mathbf{Jx}\| \geq m\|\mathbf{x}\|$ for all x. Thus, for any perturbation satisfying $\|\mathbf{J}_\epsilon\| < m$, we have $\|(\mathbf{J} + \mathbf{J}_\epsilon)\mathbf{x}\| > 0$. It follows that $\mathbf{J} + \mathbf{J}_\epsilon$ is also injective, and we conclude that the subset of injective $\mathbf{J}$ is open.

Because its range is closed, surjectivity of $\mathbf{J}$ is equivalent to injectivity of the adjoint $\mathbf{J}^*$, so the same argument on adjoints implies that the subset of surjective $\mathbf{J}$ is open. It follows that the union (the set of $\mathbf{J}$ that are injective or surjective) is also open.

Now, we consider denseness. Suppose we have some quasi-Toeplitz $\mathbf{J}$ that is neither injective nor surjective. Define $n = \min(\text{nonex}(\mathbf{J}), \text{indet}(\mathbf{J}))$. Choose orthonormal sets $e_1, \ldots, e_n$ and $f_1, \ldots, f_n$ in the kernel null $\mathbf{J}$ and the cokernel $(\text{range } \mathbf{J})^\perp$, respectively. Let $\tilde{\mathbf{E}}$ be the finite-rank operator mapping each e_i to f_i (and $(\text{null } \mathbf{J})^\perp$ to zero). We note that $\|\tilde{\mathbf{E}}\| = 1$, and that the ranges of $\mathbf{J}$ and $\tilde{\mathbf{E}}$ are orthogonal. Then:

- If $n = \text{indet}(\mathbf{J})$, then the null spaces of $\mathbf{J}$ and $\tilde{\mathbf{E}}$ are also orthogonal, and thus either $\mathbf{Jx}$ or $\tilde{\mathbf{E}}\mathbf{x}$ must be nonzero for any nonzero $\mathbf{x}$. Hence for any ϵ and nonzero $\mathbf{x}$, we have $\|(\mathbf{J} + \epsilon\tilde{\mathbf{E}})\mathbf{x}\|^2 = \|\mathbf{Jx}\|^2 + \epsilon^2\|\tilde{\mathbf{E}}\mathbf{x}\|^2 > 0$. It follows that $\mathbf{J} + \epsilon\tilde{\mathbf{E}}$ is injective.

- If $n = \text{nonex}(\mathbf{J})$, then the ranges of $\mathbf{J}$ and $\tilde{\mathbf{E}}$ provide a direct sum decomposition of ℓ^2, and hence for any ϵ, $\mathbf{J} + \epsilon\tilde{\mathbf{E}}$ is surjective.

We conclude that by choosing arbitrarily small ϵ, we can always find a $\mathbf{J} + \epsilon\tilde{\mathbf{E}}$ (which is still quasi-Toeplitz since $\tilde{\mathbf{E}}$ has finite rank) arbitrarily close to $\mathbf{J}$ that is injective or surjective.[A-8] Combining with our earlier insights, it follows that quasi-Toeplitz $\mathbf{J}$ are generically injective or surjective, and thus that proposition 6 (aside from the structure of $\mathbf{J}^{-1}$ when it exists) holds generically.

Finally, we need to show that $\mathbf{J}^{-1}$, when it exists, is quasi-Toeplitz. To do so, we note that if $\mathbf{J}$ is invertible, the winding number $\text{wind}(j)$ must be 0 by proposition 7, in which case proposition 6 tells us that $\mathbf{T}(\mathbf{j})^{-1}$ exists and is quasi-Toeplitz. We then define $\mathbf{E}_0 \equiv$

[A-8] This is a standard argument—developed, for instance, in a more general context in Theorem 2.2 of Chapter 10 in Clancey and Gohberg (1981). Onatski (2006) cites an implication of this result proven in the same text for his genericity result.

$-\mathbf{J}^{-1}\mathbf{E}T(\mathbf{j})^{-1}$, which the product of compact and bounded operators and thus compact. We can easily verify that $\mathbf{J}^{-1} = \mathbf{T}(\mathbf{j})^{-1} + \mathbf{E}_0$, and therefore $\mathbf{J}^{-1}$ is also quasi-Toeplitz.

C.4 Proof of proposition 9

The generic statements here follow from proposition 8. The fact that $\text{wind}(j) < 0$ implies indeterminacy ($\text{indet}(\mathbf{J}) > 0$) and $\text{wind}(j) > 0$ implies nonexistence ($\text{nonex}(\mathbf{J}) > 0$) follows from proposition 7 and $\text{indet}(\mathbf{J})$ and $\text{nonex}(\mathbf{J})$ both being nonnegative. The "non-generic" case, as discussed in the proof of proposition 8, is one where $\mathbf{J}$ is neither injective nor surjective, so that $\text{indet}(\mathbf{J}), \text{nonex}(\mathbf{J}) > 0$. Finally, the final part of the proof of proposition 8 showed that $\mathbf{J}^{-1}$ is quasi-Toeplitz whenever it exists.

C.5 Proofs of propositions 10 and 11

The proofs of lemma 1 and propositions 1–3 go through unchanged, replacing scalar Toeplitz entries with blocks at each step.

To extend proposition 7, with $\text{wind}(\det j)$ replacing $\text{wind}(j)$, we apply theorem 6.5 from Böttcher and Silbermann (2012), which states that block Toeplitz operators have a Fredholm index equal to $-\text{wind}(\det j)$. The same argument for genericity as in the proof of proposition 8 then goes through, and the argument for nearly all of proposition 9 is also unchanged.

All that remains is to show that when a block quasi-Toeplitz operator is invertible, its inverse is also block quasi-Toeplitz. To do so for block Toeplitz operators, we apply theorem 6.6 from Böttcher and Silbermann (2012), which provides a factorization of the symbol that implies a factorization of $T(\mathbf{j})$ into $T(\mathbf{j}_-)$ and $T(\mathbf{j}_+)$, exactly as in our proof of proposition 6. (This factorization only exists if all the "right partial indices" are zero, but this must be true for the operator to be invertible in the first place.) This, in turn, implies a block quasi-Toeplitz inverse through the same logic as in that proposition. The extension to the block *quasi*-Toeplitz case is then the same as in proposition 9.

D Appendix to section 4

D.1 Proof of proposition 12

For quasi-Toeplitz operators, this follows from the collective results in section 2 of Böttcher and Silbermann (2012). Their theorem 2.16 shows that what they call "stability" of a sequence is unaffected by adding a compact operator, if its limit is invertible. This allows us to restrict ourselves to the Toeplitz case. Then, theorem 2.11 shows that the sequence of finite sections of an invertible Toeplitz operator is stable, and proposition 2.4 shows the stability plus invertibility of the limit implies that a sequence is an "applicable" "approximation method" to its limit (as defined prior to theorem 2.1), which includes as part of its definition our desired convergence.

For block quasi-Toeplitz operators, we use the extension of these results to the block case in section 6 of the same text. In particular, theorem 6.10 says that block quasi-Toeplitz finite sections are stable if the operator is invertible *and* the block Toeplitz operator whose sequence is flipped relative to the Toeplitz part (i.e. the Toeplitz with symbol $a(z^{-1})$ vs. the original symbol $a(z)$) is also invertible. Since $a(z^{-1})$ and $a(z)$ have the same winding number, this latter requirement holds generically whenever the block quasi-Toeplitz operator is itself invertible (since this requires a winding number of zero). Section 6.2 of the text says that "propositions 2.2 and 2.4 remain valid in the block case", from which the rest of the argument goes through.

D.2 Application to quantitative IKC model

To see how generally we can obtain accurate solutions with truncated and low-rank approximations to the compact correction $\mathbf{E}$, we now take the "quantitative" version of the intertemporal Keynesian cross model from section 7 of Auclert et al. (2024a). This model has a household block with both liquid and illiquid accounts, which allows it to match both high aggregate assets and high MPCs. (In this paper's main calibration, we use discount factor heterogeneity with a single account instead.) It combines this household block with both price and wage rigidity, investment subject to capital adjustment costs, and potentially time-varying profits that are capitalized into traded assets.

We use the exact same calibration as in Auclert et al. (2024a), including for the fiscal rule parameter $\rho_B = 0.93$, and leave the details to that paper. We then ask the following question: if we compute the "general equilibrium Jacobian" $\mathcal{G}$ that maps shocks $d\mathbf{G}$ to government spending to the resulting changes $d\mathbf{Y}$ in output, and then replace it with a quasi-Toeplitz Jacobian that uses an approximation to its compact correction $\mathbf{E}$, truncated to $T \times T$ and compressed to rank k, what is the max absolute error in $d\mathbf{Y}$? We do this for a shock $dG_t = 0.76^t$, which is the baseline from Auclert et al. (2024a).

Figure D.1 plots the results, analogously to figure 12. Due to the internal persistence of the model, we cannot truncate $\mathbf{E}$ too aggressively, but a low-rank approximation is still quite accurate: for instance, truncating to $T = 1000$ and using rank 10, we get accuracy to 7 digits in the path of $d\mathbf{Y}$. Truncating to $T = 1500$ and using rank 15, we get accuracy to 9

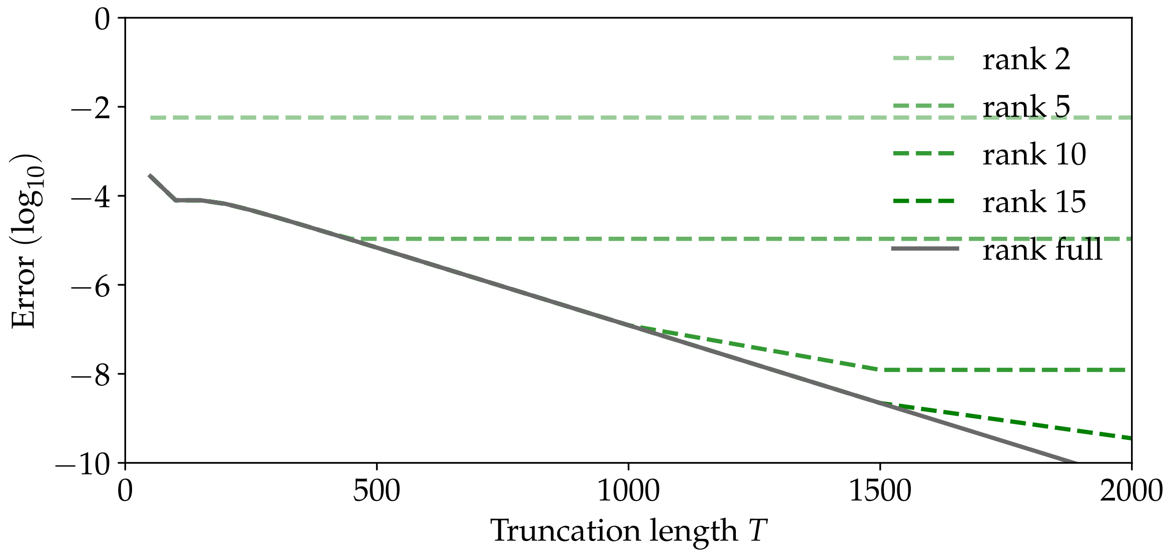


Figure D.1: Error in $d\mathbf{Y} = \mathcal{G}d\mathbf{G}$ when approximating correction to $\mathcal{G}$

digits.

D.3 Application to menu cost models

For another test of whether heterogeneous-agent Jacobians can be accurately represented with low-rank approximations to their corrections, we now turn to menu cost models, following Auclert et al. (2024c). As the paper describes, there are two key Jacobians of interest: the "pass-through matrix" Ψ, which maps aggregate log nominal marginal cost shocks to aggregate log price changes, and the "generalized Phillips curve" $\mathbf{K}$, which maps aggregate log real marginal cost changes to aggregate inflation changes. For each Jacobian's compact correction, we will describe as the "approximate rank" the number of singular values that are greater than 10^{-10}.[A-9]

To start, we use the pure menu-cost calibration in that paper (which it calls "Golosov-Lucas"), with a quarterly price adjustment frequency of 23.9%.[A-10] For this calibration, the pass-through matrix Ψ has a correction with approximate rank 2, and the generalized Phillips curve $\mathbf{K}$ has a correction with approximate rank 3.

[A-9] Here we focus on approximate rank rather than also studying truncation length as in figure D.1. This is because for pass-through matrices, all interaction is very local in time (making truncation mostly irrelevant), while the generalized Phillips curve is extremely forward-looking (making truncation a major source of inaccuracy that would swamp the effects of low rank).

[A-10] Given the simple structure of the model, with normal innovations to single "log price gap" state variable, the price adjustment frequency—which varies inversely with the size of the Ss bands relative to the innovations—fully characterizes the calibration, aside from the discount factor β, which we set to 0.99.

Calibration	Effective rank of correction	
	Pass-through Ψ	GPC $\mathbf{K}$
Freq = 23.9%, no free resets	2	3
Freq = 23.9%, 75% free resets	3	3
Freq = 5%, 75% free resets	5	5

Table D.1: Summary of results for menu cost models

Next, we use a calibration (called "Nakamura-Steinsson" in the paper) where some price changes are free resets, which arrive with some Calvo frequency. This is calibrated to the same overall quarterly price adjustment frequency of 23.9%, but targeting 75% of adjustments being "free resets" rather than involving a menu cost. For this calibration, both the pass-through matrix Ψ and the generalized Phillips curve $\mathbf{K}$ have corrections with approximate rank 3.

Finally, to test the limits of this result, we consider a calibration where 75% of price adjustments are still free, but with a price adjustment frequency of 5%—since Auclert et al. (2024c) generally find slightly more complex behavior in models with lower price adjustment frequencies. (We can think of this as the same model but with shorter periods.) For this calibration, the pass-through matrix Ψ and generalized Phillips curve $\mathbf{K}$ both have corrections with approximate rank 5.

These results are summarized in table D.1.

E Appendix to section 5

E.1 Solving the global system

Getting the last equation. Stacking the equations given by (35) side-by-side yields

$$d\mathbf{Y}^m = \mathbf{M}(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' + d\mathbf{G}^m, \tag{A.32}$$

where superscript m denotes the $T \times N$ matrix of country-specific responses. As with $\mathbf{I} - \mathbf{M}$, a dimension is missing from the range of (A.32), making it numerically problematic to solve (A.32) directly. We instead replace the T goods market clearing conditions for one country in (A.32) with T consolidated international asset market clearing conditions. In matrix form, these conditions are

$$\mathbf{A}(d\mathbf{Y}^m - d\mathbf{T}^m)\mathbf{1}_{N\times 1} = d\mathbf{B}^m\mathbf{1}_{N\times 1}. \tag{A.33}$$

We can derive (A.33) from (A.32) by first subtracting $d\mathbf{T}^m$ from both sides and splitting up the first term of the right-hand side:

$$d\mathbf{Y}^m - d\mathbf{T}^m = \mathbf{M}(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' + (d\mathbf{G}^m - d\mathbf{T}^m) \tag{A.34}$$

$$= (\mathbf{M} - \mathbf{I}_T)(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' + \mathbf{I}_T(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' + (d\mathbf{G}^m - d\mathbf{T}^m). \tag{A.35}$$

Then, gathering terms, we obtain

$$(\mathbf{I}_T - \mathbf{M})(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' = (d\mathbf{Y}^m - d\mathbf{T}^m)(\Pi' - \mathbf{I}_N) + (d\mathbf{G}^m - d\mathbf{T}^m). \tag{A.36}$$

Pre-multiplying by $\mathbf{K}$ (the operator from Auclert et al. (2024a) consistent with $\mathbf{A} = \mathbf{K}(\mathbf{I} - \mathbf{M})$) yields

$$\mathbf{A}(d\mathbf{Y}^m - d\mathbf{T}^m)\Pi' = \mathbf{K}(d\mathbf{Y}^m - d\mathbf{T}^m)(\Pi' - \mathbf{I}_N) + d\mathbf{B}^m. \tag{A.37}$$

Post-multiplying by $\mathbf{1}_{N\times 1}$, we find (A.33), since $(\Pi' - \mathbf{I}_N)\mathbf{1}_{N\times 1} = \mathbf{0}_{N\times 1}$.

Setting up the full system. In vectorized form, (A.32) may be rearranged into

$$\underbrace{(\mathbf{I}_{NT} - \Pi \otimes \mathbf{M})}_{\equiv \mathbf{H}} d\mathbf{Y}^v = \underbrace{d\mathbf{G}^v - (\Pi \otimes \mathbf{M})d\mathbf{T}^v}_{\equiv \mathbf{b}}, \tag{A.38}$$

where superscript v denotes the the vectorized form of any variable with an m superscript. $\mathbf{H}$ is a block matrix with N^2 $T \times T$ blocks. We then replace the first block row of $\mathbf{H}$ ($\mathbf{H}_{(1:T,:)}$ in Matlab index notation) with

$$(\mathbf{1}_{1\times N} \otimes \mathbf{A})d\mathbf{Y}^v \tag{A.39}$$

and the first T entries of $\mathbf{b}$ ($\mathbf{b}_{1:T}$ in Matlab index notation) with

$$d\mathbf{B}^m\mathbf{1}_{N\times 1} + (\mathbf{1}_{1\times N} \otimes \mathbf{A})d\mathbf{T}^v. \tag{A.40}$$

Denote the matrix $\mathbf{H}$ with the first block row replaced by $\widetilde{\mathbf{H}}$ and, likewise, define $\widetilde{\mathbf{b}}$ as the vector $\mathbf{b}$ with the first T entries replaced. Hence, the solvable system is

$$\widetilde{\mathbf{H}} d\mathbf{Y}^v = \widetilde{\mathbf{b}}. \tag{A.41}$$

One step remains: to turn (A.41) into the system we will ultimately solve. $\widetilde{\mathbf{H}}$ is a block matrix with N^2 $T \times T$ blocks, but we can rearrange it into a matrix $\widehat{\mathbf{H}}$ with T^2 $N \times N$ blocks instead. Let $d\widehat{\mathbf{Y}}^v$ and $\widehat{\mathbf{b}}$ be the appropriately rearranged versions of $d\mathbf{Y}^v$ and $\widetilde{\mathbf{b}}$, as well. The final system is then

$$\widehat{\mathbf{H}} d\widehat{\mathbf{Y}}^v = \widehat{\mathbf{b}}. \tag{A.42}$$

One note: despite this notation, it is not necessary to build $\widehat{\mathbf{H}}$ explicitly. Building the full matrix in memory is inefficient, and often prohibitively costly from a memory standpoint. It is better to directly evaluate $\widehat{\mathbf{H}}$ times a vector, taking advantage of the structure of $\widehat{\mathbf{H}}$.

Solving (A.42). To solve this model, we follow the guidance of sections 4.2 and 4.3 and use preconditioned GMRES. Because $\widehat{\mathbf{H}}$ is block quasi-Toeplitz $T(\widehat{\mathbf{h}}) + \mathbf{E}$, we use $T(\widehat{\mathbf{h}}^{-1})$ as the pre-conditioner. The symbol $\widehat{\mathbf{h}}$ of the block-Toeplitz part of $\widehat{\mathbf{H}}$ can be obtained straightforwardly following section A.2 and its inverse $\widehat{\mathbf{h}}^{-1}$ can be computed using the methods in section A.3.

Having the inverse symbol $\widehat{\mathbf{h}}^{-1}$, we define a function to apply $T(\widehat{\mathbf{h}}^{-1})$ to a vector using efficient FFT methods (following A.3). Then, all that remains is to produce a function to compute $\widehat{\mathbf{H}}\mathbf{x}$ (for any $\mathbf{x}$) using pieces of (A.32) and (A.33), both of which can be evaluated cheaply. Those two functions are supplied to a GMRES solver along with an initial guess and an error tolerance.

E.2 Global effect of the tax-financed fiscal shock

Figures E.1 and E.2 plot the output effects of our fiscal shock—a deficit-financed tax cut in the US at date 0, which after roughly 10 quarters starts being paid down—for all countries in the world, as a percentage of their steady-state GDP. On impact, figure E.1 shows that the output boom is overwhelmingly concentrated in the US (and to a lesser degree in Canada and Mexico). After 20 quarters, figure E.2 shows that output has returned roughly to steady state in the US—as the contractionary effects of paying down the debt begin to hit—but there is still a slight boom in Canada and Mexico, and an even slighter boom in the rest of the world (note the different scale from the previous figure).

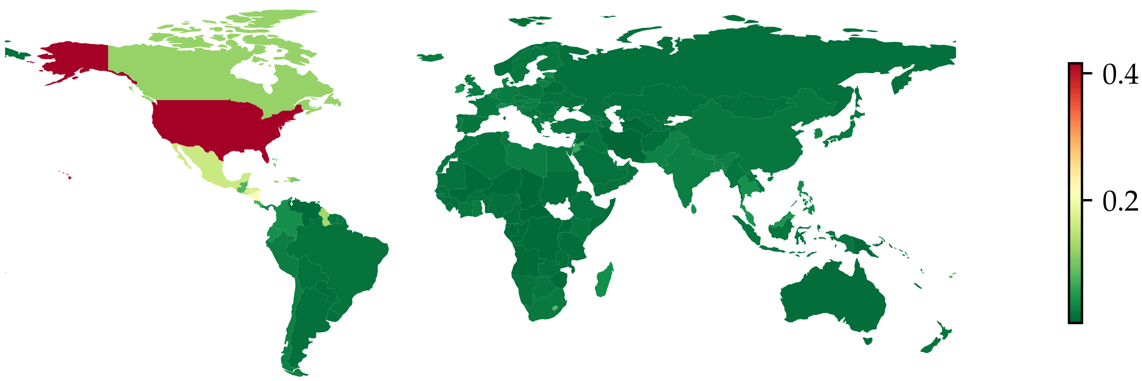


Figure E.1: Impact effect of deficit-financed US tax cut on output, as percent of GDP

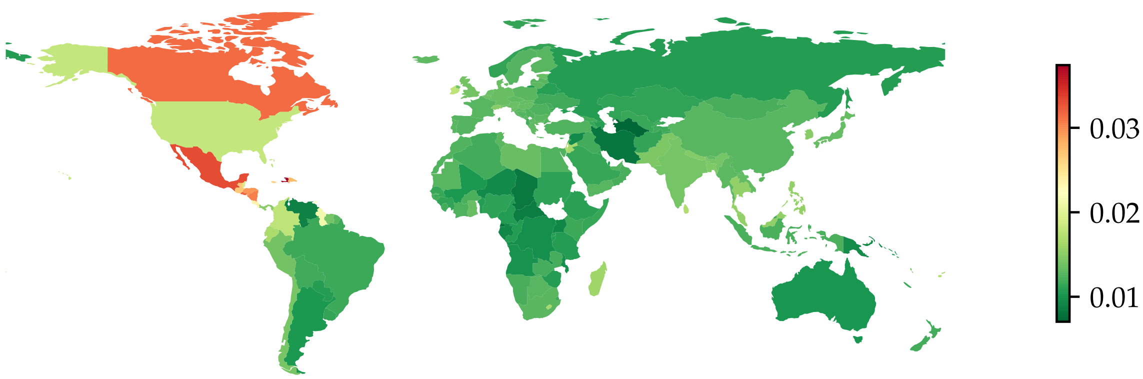


Figure E.2: Effect of US fiscal shock on output after 20 quarters, as percent of GDP