Thinking Big:
Determinacy and Large-Scale Solutions in the Sequence Space

Adrien Auclert Stanford
Evan Majic Northwestern
Matt Rognlie Northwestern
Ludwig Straub Harvard

Solving het-agent models to first order

- Two key considerations:
  - Size of idiosyncratic state space **S**
  - Number of endogenous aggregate variables **N**
- **State-space approach:** costly when S large. Has determinacy criterion.

[Reiter, Ahn-Kaplan-Moll-Winberry-Wolf, Bayer-Luetticke, …]

- **Sequence-space approach:** fast when S large, costly when N large.

[Boppart-Krusell-Mitman, Auclert-Bardoczy-Rognlie-Straub, ...]

- How do we solve models <u>when both S and N are large?</u>

Introducing... SSJ 2.0!

- Obtain a **structure theorem** for sequence-space Jacobians
- When het-agent model is *stationary*, Jacobians are *quasi-Toeplitz*:

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

i.e. sum of a Toeplitz operator $\mathbf{T}(\mathbf{j})$ and a compact operator $\mathbf{E}$

(When represented as matrix, "Toeplitz" means "each diagonal is constant.)

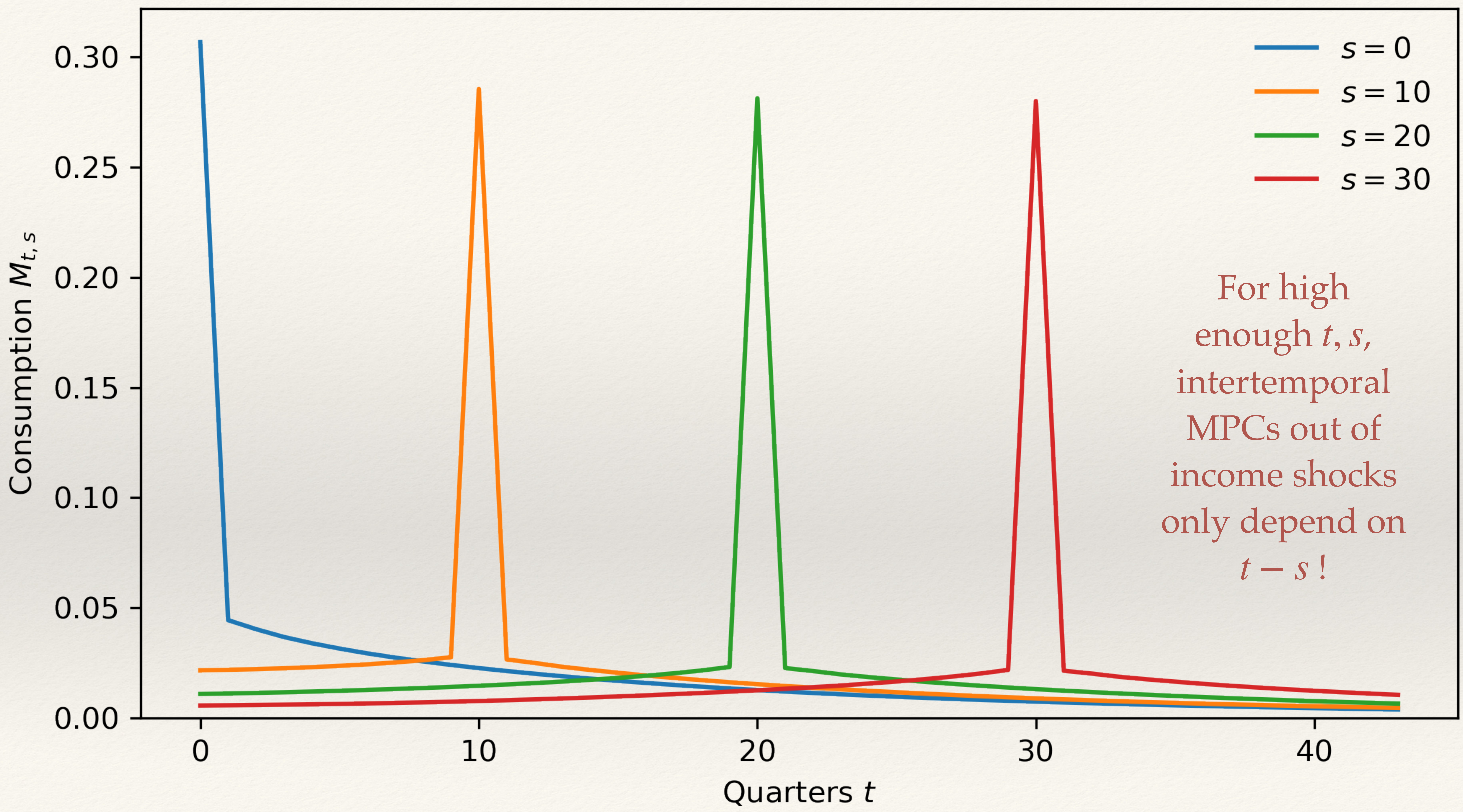
- Exploit this structure (generalizing to $N > 1$ case) in many ways:
  - **Winding number criterion** on $\mathbf{j}$ for **determinacy** & existence
  - More accurate computations working directly with $\mathbf{j}$ and $\mathbf{E}$
  - Using $\mathbf{T}(\mathbf{j}^{-1})$ as guess for $\mathbf{J}^{-1}$ gives **rapid iterative solution**, even when N large

Structure theorem

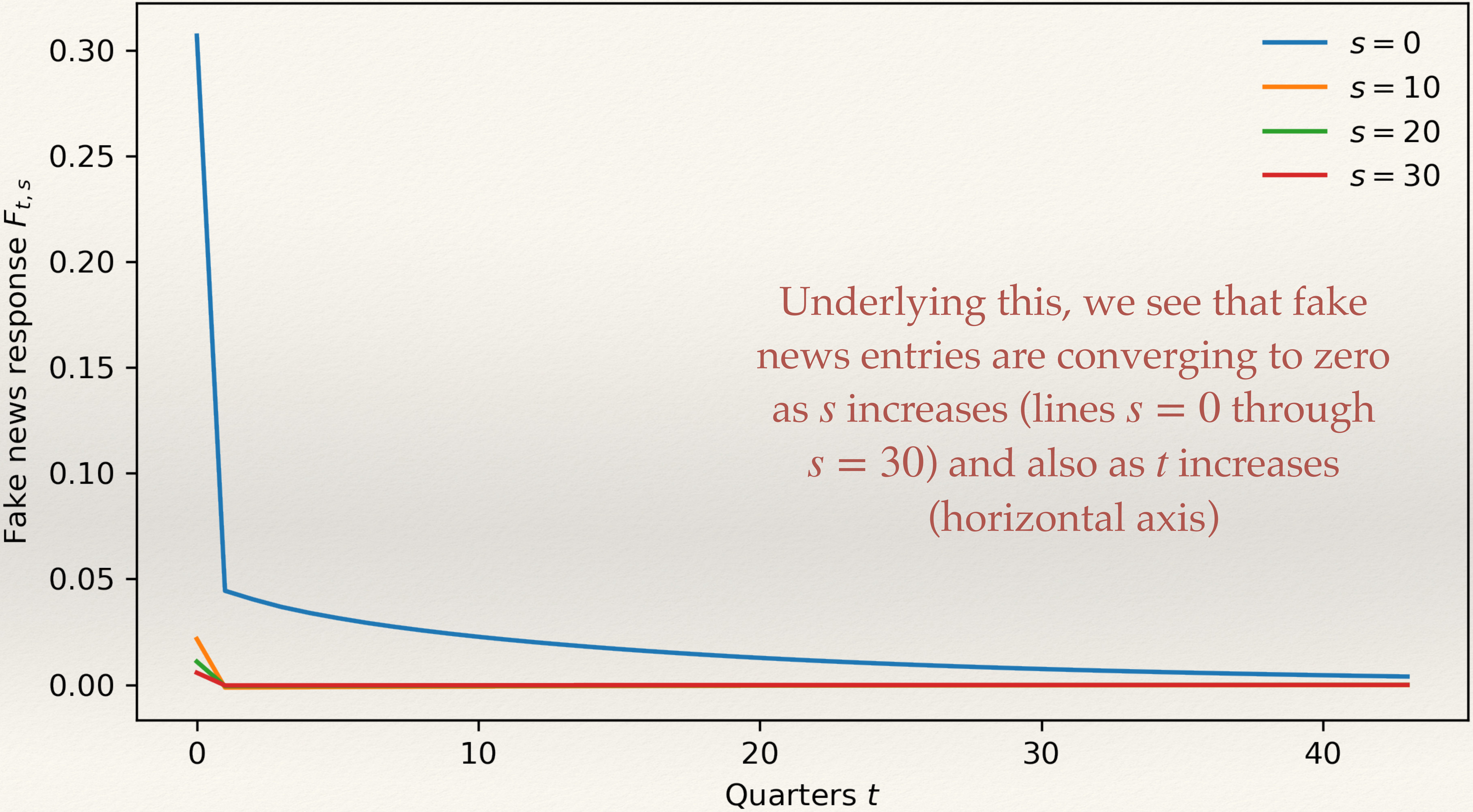
Building block: the fake news matrix

- Recall that Jacobian is cumulative diagonal sum of fake news matrix, i.e. $J_{2,3} = F_{0,1} + F_{1,2} + F_{2,3}$
- If $F_{t,s} \to 0$ fast enough as $t, s \to \infty$, then each diagonal of $\mathbf{J}$ will converge to a constant j_{t-s}
- Interpretation: if shock anticipated far enough in advance, only position $t-s$ vs. shock matters

$$\begin{pmatrix} F_{00} & F_{01} & F_{02} & \cdots \\ F_{10} & F_{11} & F_{12} & \cdots \\ F_{20} & F_{21} & F_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$



For high enough t, s, intertemporal MPCs out of income shocks only depend on $t - s$!



Underlying this, we see that fake news entries are converging to zero as s increases (lines $s = 0$ through $s = 30$) and also as t increases (horizontal axis)

Why would fake news entries converge to zero?

- $F_{t,s}$ is the effect at date t of having thought at date 0 that there would be shock at date s
- Plausible that:
  - Effect on date-0 policy decays as horizon s increases
  - Persistent effect from date-0 policy decays as t increases
- If both, we say model is **stationary**

$$\begin{pmatrix} F_{00} & F_{01} & F_{02} & \cdots \\ F_{10} & F_{11} & F_{12} & \cdots \\ F_{20} & F_{21} & F_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Fake news matrix of stationary het-agent models

- For $t > 0$, SSJ paper shows that $F_{t,s}$ takes the form:

$$F_{t,s} = \mathscr{E}_{t-1}' \mathscr{D}_s$$

  - $\mathscr{D}_s$ is effect of anticipating date-s shock at date 0 on the date-1 distribution
  - $\mathscr{E}_{t-1}$ is the *expectation function*, giving expected quantity of interest (e.g. consumption, assets) in $t-1$ periods, if steady-state policy followed

- Eventual decay at rate of at least $\mathscr{D}_s \sim \beta^s$ from discounting as $s \to \infty$

(Note $\mathbf{1}'\mathscr{D}_s = 0$, since distribution mass preserved.)

- If Λ_{ss} ergodic, $\mathscr{E}_t$ approaches multiple of $\mathbf{1}$ at rate γ^t, with γ largest eigenvalue < 1
- If both: $|F_{t,s}| \leq K\beta^s\gamma^t \leq K\Delta^{s+t}$, for $\Delta \equiv \max(\beta, \gamma) < 1$, **we call model "stationary"**

Using this to characterize Jacobian

- If $|F_{t,s}| \leq K\Delta^{s+t}$, then we can define and guarantee convergence for:

$$j_u = F_{u,0} + F_{u+1,1} + F_{u+2,2} + \dots$$

i.e. the sum of all entries in **F** on the uth lower diagonal (analogous for $u < 0$)

- We then have $J_{t+i,s+i} \to j_{t-s}$ as $i \to \infty$ for any t, s [“asymptotically Toeplitz”]
- j_{t-s} is response $t - s$ periods after shock, if shock infinitely well-anticipated
- We also have $J_{t,s} - j_{t-s} \equiv E_{t,s} = -(F_{t+1,s+1} + F_{t+2,s+2} + \dots) \leq \dfrac{K}{1-\Delta^2}\Delta^{t+s}$

[E is “correction” vs. exact Toeplitz, with terms on diagonal not yet summed]

Quasi-Toeplitz form

- Summing up: for any Jacobian $\mathbf{J}$ of a stationary het-agent model, we have $J_{t,s} - j_{t-s} = E_{t,s}$, where $|E_{t,s}|$ bounded by multiple of Δ^{t+s}
- Can write as sum of Toeplitz operator $T(\mathbf{j})$ and **compact** "correction" operator $\mathbf{E}$:

$$\mathbf{J} = \underbrace{\begin{pmatrix} j_0 & j_{-1} & j_{-2} & \\ j_1 & j_0 & j_{-1} & \ddots \\ j_2 & j_1 & j_0 & \ddots \\ & \ddots & \ddots & \ddots \end{pmatrix}}_{\equiv T(\mathbf{j})} + \mathbf{E}$$

("Compact" on ℓ^2 means limit of finite-rank operators, behaves similarly to a finite-dimensional matrix. $|E_{t,s}| \leq K'\Delta^{t+s}$ readily implies this.)

- This is called a **quasi-Toeplitz operator**, and has many nice properties!

Why might this representation be useful?

- Quasi-Toeplitz operators are **closed** under addition, multiplication, etc., and even inversion, assuming an inverse exists [e.g. Bini, Massei, Robol 2019]
  - so we can chain along DAG, solve for unknowns, and stay quasi-Toeplitz!
  - (simple aggregate equations already have Toeplitz Jacobians)
- Toeplitz has nice theory for **existence & uniqueness** of solutions; this mostly extends to quasi-Toeplitz
- In practice, $\mathbf{E}$ often very well-approximated by **low-rank** matrix
  - So can represent and work with $\mathbf{J}$ a lot more efficiently than ordinary $T \times T$

Existence and uniqueness of solutions

The winding number

- Recall: j_k is the entry on the kth lower diagonal of Toeplitz $T(\mathbf{j})$
- $\{j_k\}_{k=-\infty}^{\infty}$ is a two-sided sequence, and we say its **symbol** is the Laurent series

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

- The **winding number** $\mathrm{wind}(j)$ is # of times $j(z)$ rotates counterclockwise around 0 as z goes counterclockwise around the unit circle
- Standard result: $T(\mathbf{j})$ is invertible iff its winding number is zero!
  - If $\mathrm{wind}(j) < 0$, then surjective but not injective; vice versa if $\mathrm{wind}(j) < 0$

Simple examples of winding number

- Lag operator $\mathbf{L} : (x_0, x_1, \ldots) \mapsto (0, x_0, x_1, \ldots)$ is injective but not surjective
  - symbol $j(z) = z$, with winding number 1 (j goes counterclockwise once)
- Lead operator $\mathbf{F} : (x_0, x_1, \ldots) \mapsto (x_1, x_2, \ldots)$ is surjective but not injective
  - symbol $j(z) = z^{-1}$, with winding number -1 (j goes clockwise instead)

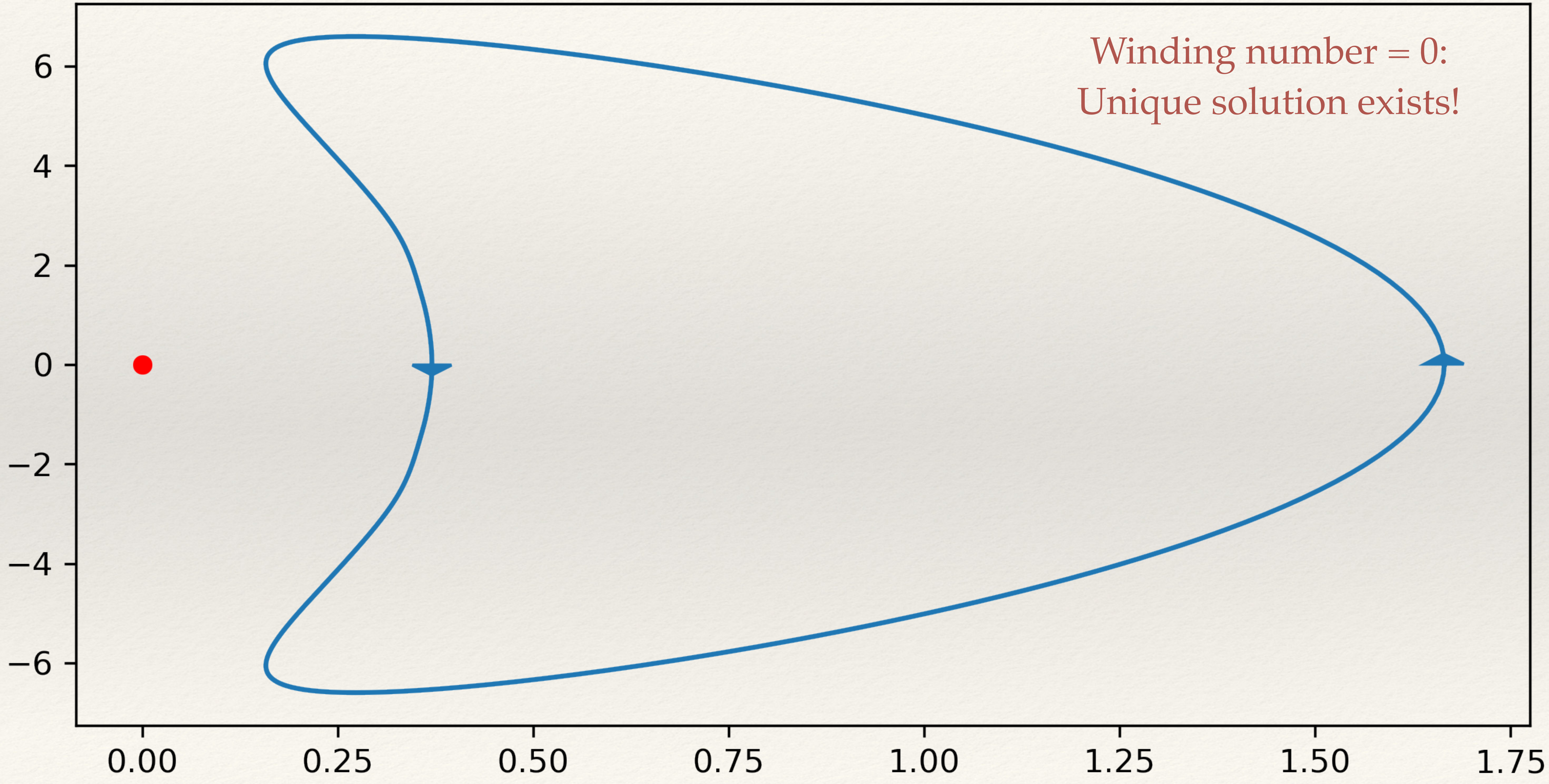
- More general result: winding number $n > 0$ implies n dimensions missing from range, $n < 0$ implies null-space of dimension $-n$ [consider examples $\mathbf{L}^n$ and $\mathbf{F}^n$]
  - loosely, winding number of n says "this Toeplitz similar to taking n lags"

Extending to quasi-Toeplitz

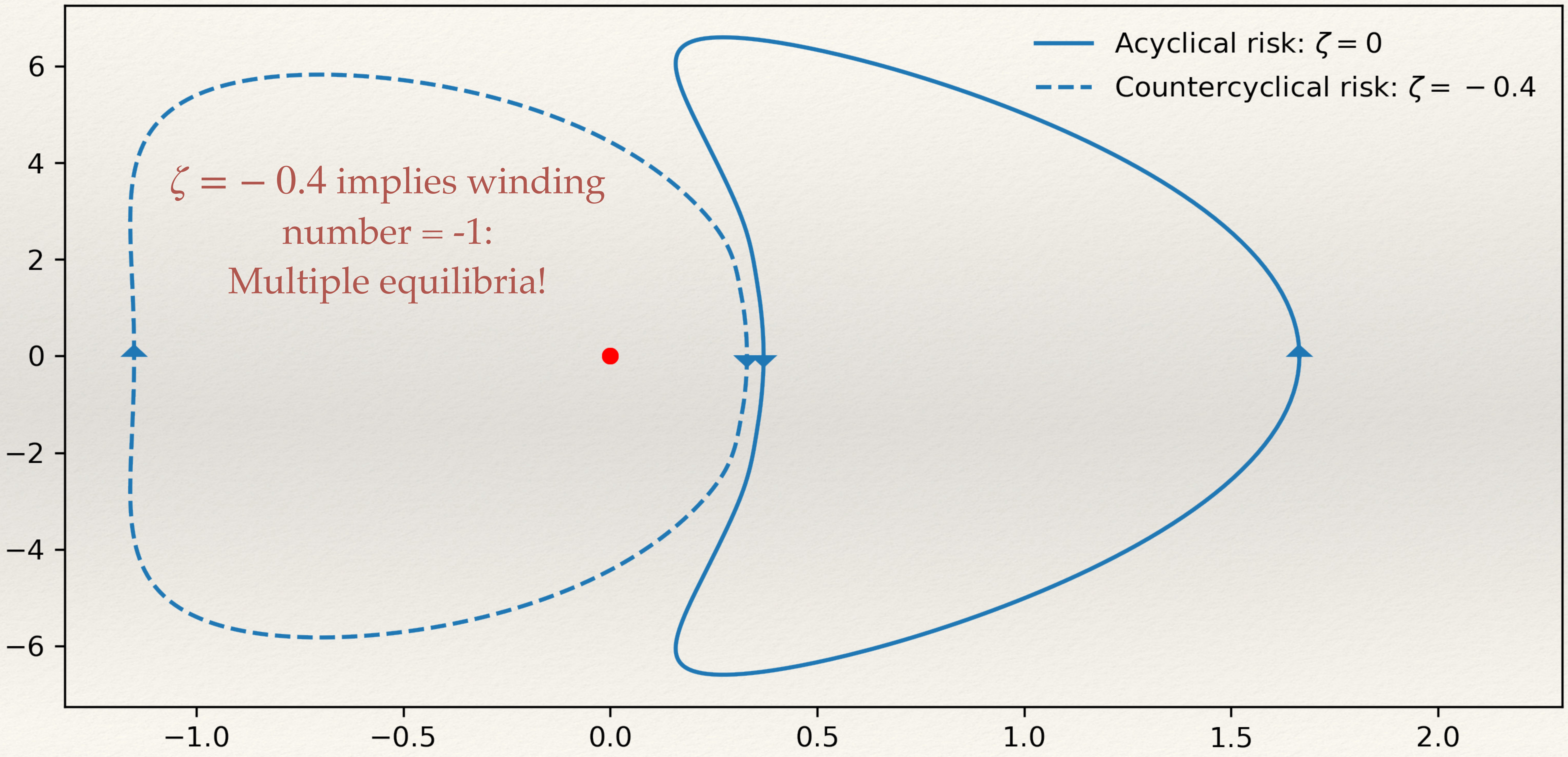
- Same result doesn't always hold for quasi-Toeplitz, but holds "generically" on an open and dense set of $\mathbf{E}$ (i.e. almost all $\mathbf{E}$):
  - If $\text{wind}(j) = 0$, then $\mathbf{J}$ is generically invertible
  - If $\text{wind}(j) < 0$, then $\mathbf{J}$ is not injective, but generically surjective
  - If $\text{wind}(j) > 0$, then $\mathbf{J}$ is not surjective, but generically injective

- Solving asset market IKC $\mathbf{A}(d\mathbf{Y} - d\mathbf{T}) = d\mathbf{B}$, then (generically) exists unique solution if $\text{wind}(a) = 0$, indeterminacy if $\text{wind}(a) < 0$, nonexistence if > 0

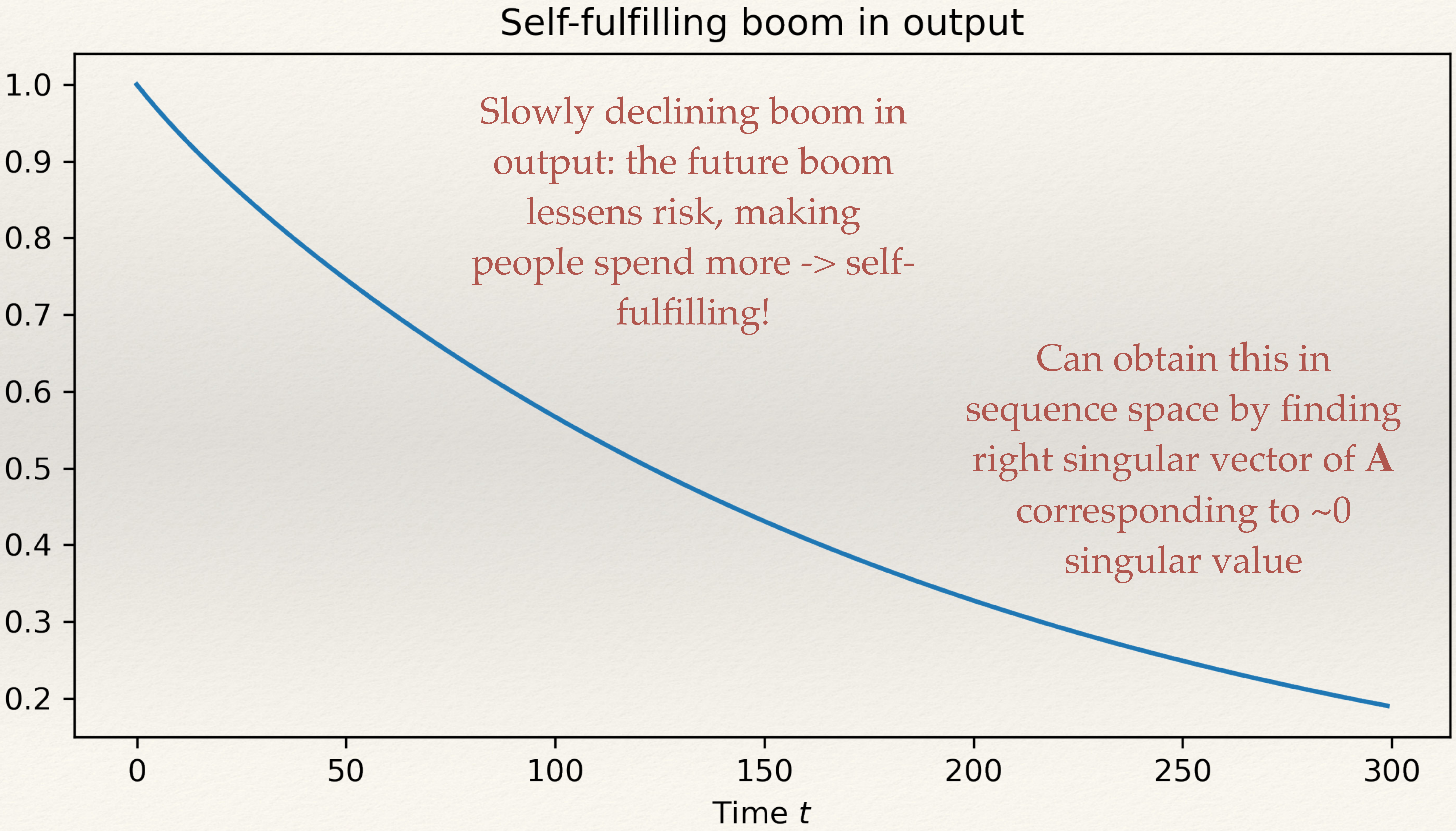
Winding number plot for **A** with standard calibration



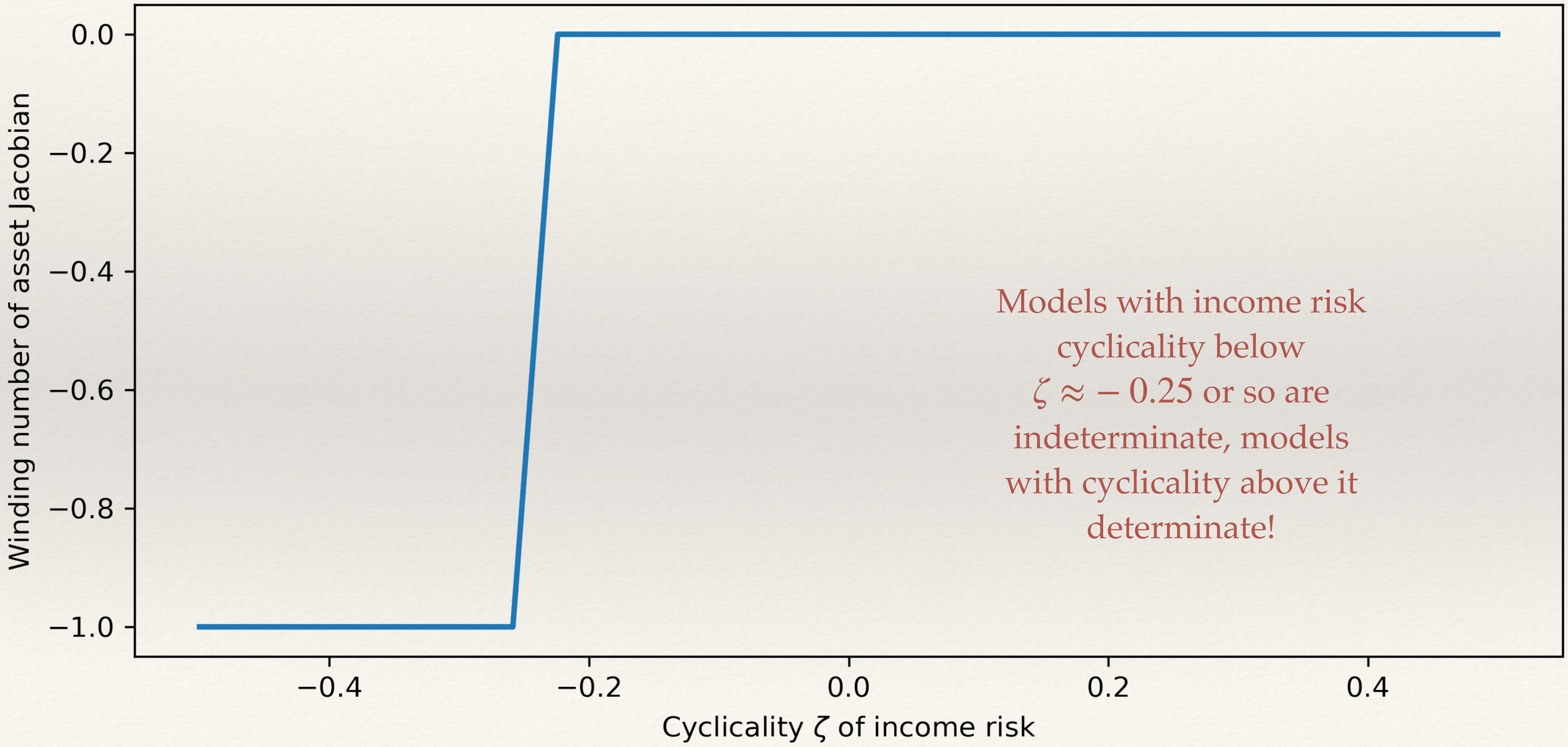
Compare to case with countercyclical risk



Can find shape of multiplicity in sequence space



How does winding number vary with ζ in general?



Block quasi-Toeplitz case

- Say we have N^2 quasi-Toeplitz matrices from N unknowns to one of N targets
- Can think of this as being one **block quasi-Toeplitz operator**, like a quasi-Toeplitz but where entries are each $N \times N$ blocks
- Then $\{j_k\}_{k=-\infty}^{\infty}$ is two-sided sequence of $N \times N$ matrices, so *matrix-valued* $j(z)$:

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

- Winding number test still holds generically, now for wind$(\det j)$
- Important case in practice, we're still working out details [see also Onatski 2006]

Operations with quasi-Toeplitz operators: No more truncation!

Directly use quasi-Toeplitz form

- We have quasi-Toeplitz representation $\mathbf{J} = T(\mathbf{j}) + \mathbf{E}$ of Jacobians
- So far, we've used the winding number of $\mathbf{j}$ to assess determinacy
- Another benefit: use this special form directly to do computations!
    - key supporting fact: in practice $\mathbf{E}$ well-approximated by low rank
    - more efficient, no longer fully truncating at T ($T(\mathbf{j})$ in principle infinite)

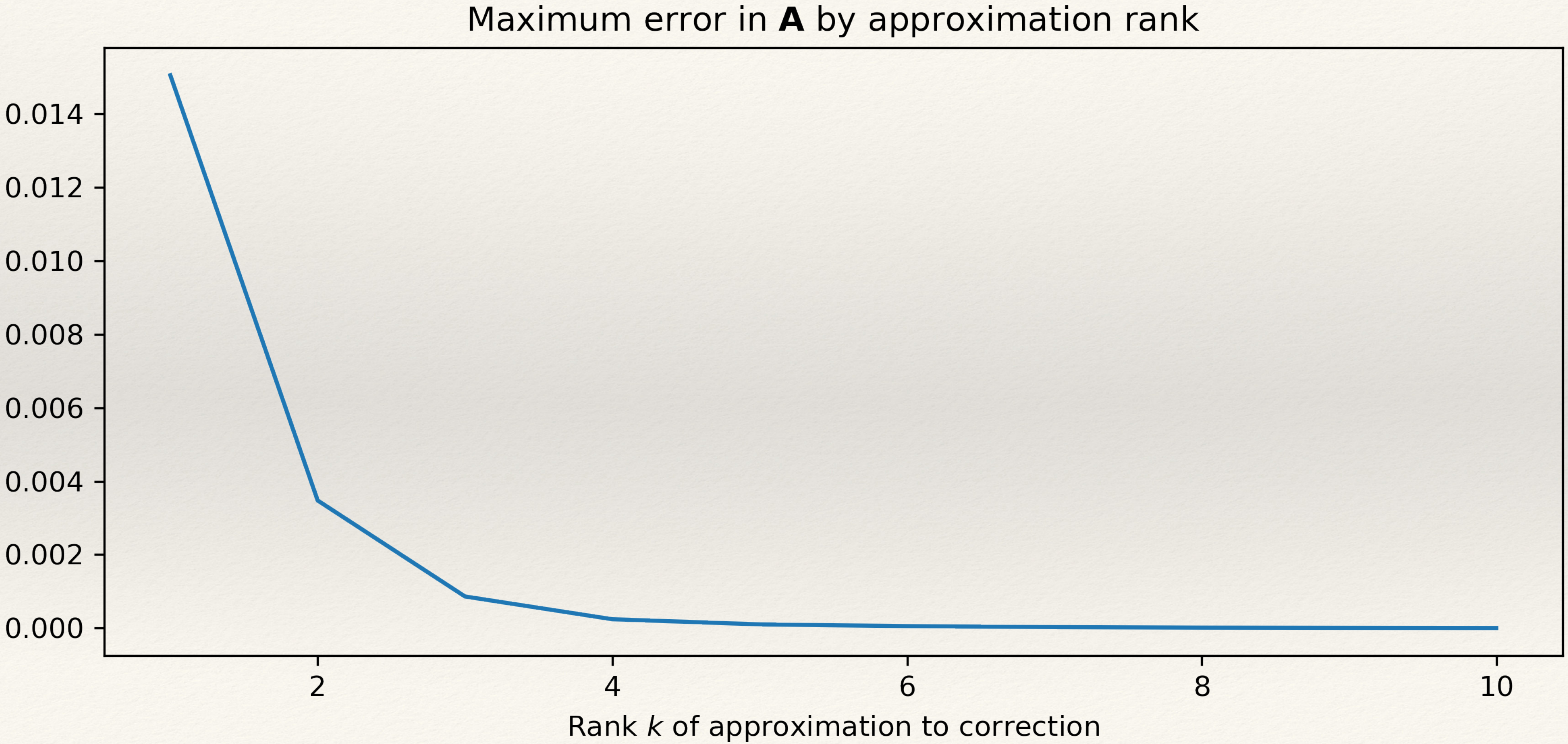
- Our inspiration: Bini, Massei, Robol (2019), "Quasi-Toeplitz matrix arithmetic: a MATLAB toolbox", paper from applied math literature

Why is this math so nice?

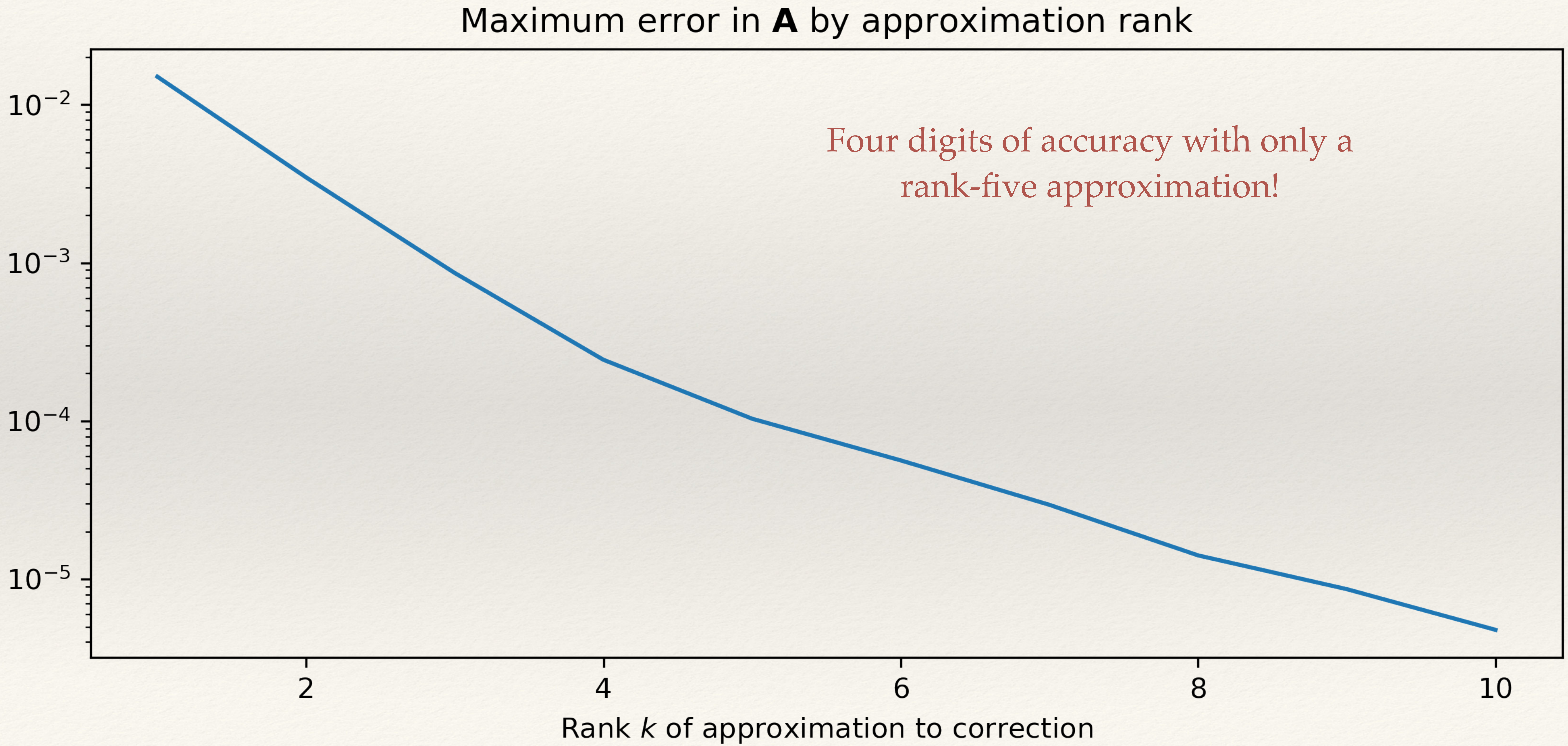
- Suppose we can write $\mathbf{J} = T(\mathbf{j}) + \mathbf{UV}'$, where $\mathbf{U}, \mathbf{V}$ are $n \times k$ matrices
  - if k low, we have a *low-rank approximation* of $\mathbf{E} \approx \mathbf{UV}'$
  - represents $\mathbf{J}$ in a more concise way

- Can define algebra of operations on $\mathbf{j}, \mathbf{U}, \mathbf{V}$
  - e.g. multiplying $\mathbf{J}_1$ and $\mathbf{J}_2$ involves multiplying $T(\mathbf{j}_1)$ and $T(\mathbf{j}_2)$, which produces quasi-Toeplitz of form $T(\mathbf{j}_1\mathbf{j}_2) + \mathbf{UV}'$; and also $T(\mathbf{j}_1)$ times $\mathbf{U}_2$, etc.
  - similar, though a bit more complicated, for inversion

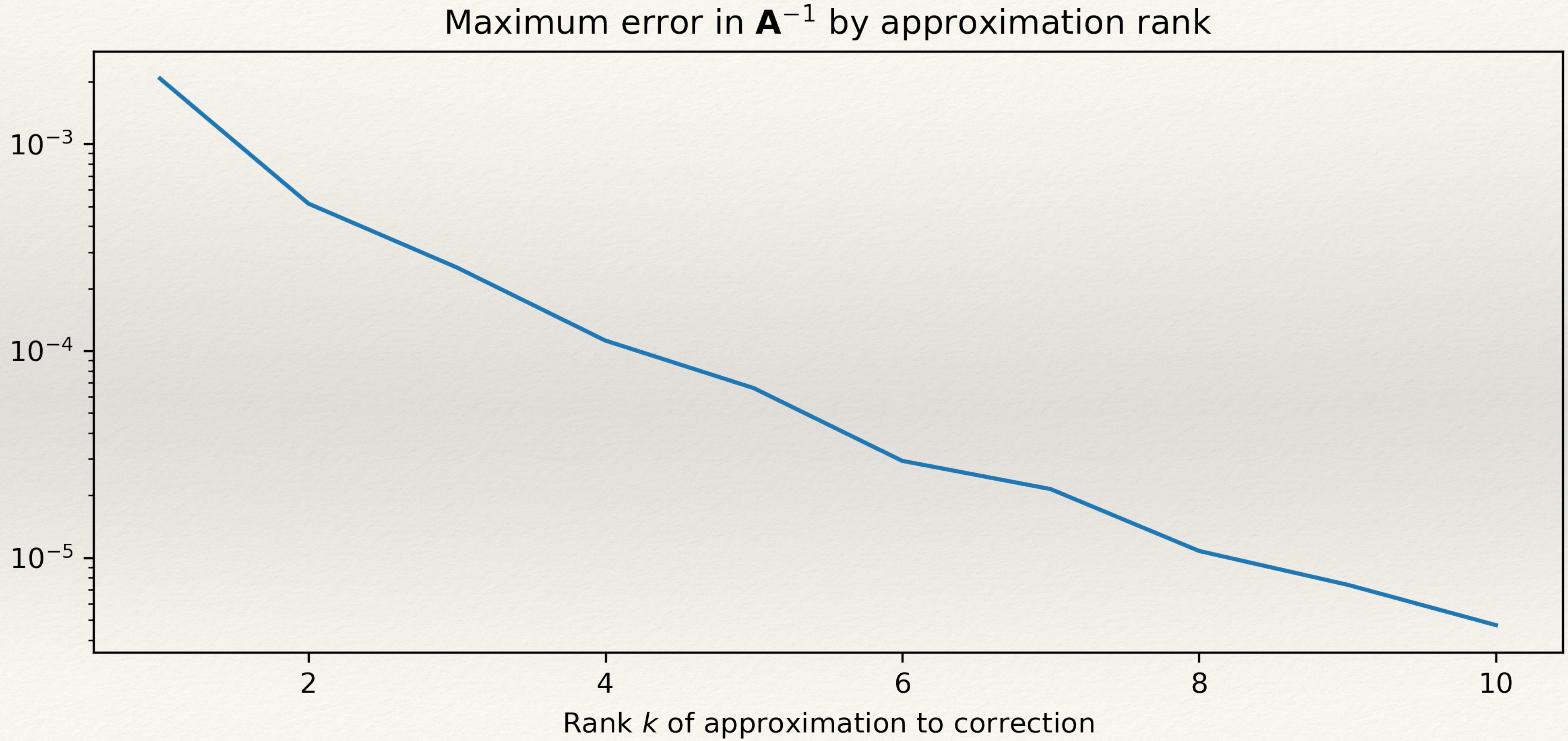
How well can we approximate $\mathbf{A}$?



Big easier to visualize with a log scale…



Low-rank approximation to $\mathbf{A}^{-1}$ also close!



How “compressed” is this?

- Suppose that $T = 1000$, so that $\mathbf{A}$ is 1000×1000, with 1 million entries!
- Can store corresponding $\mathbf{a}$ of length only ~ 2000
- Then $\mathbf{U}$, $\mathbf{V}$ in a rank-five approximation to $\mathbf{E}$ each size 5000
- So, can store near-exact approximation with only 12,000 numbers!
  - Compression of more than 80x
  - Actually can do **far better**, since many entries in $\mathbf{a}$, $\mathbf{U}$, $\mathbf{V}$ near zero
  - Doing math with this will be more accurate than with truncated $\mathbf{A}$, since no error from artificial T [implementation details too much for today, though!]

Alternative: use structure for iterative solutions (and solve giant models in the process!)

First point: easy to get Toeplitz part of inverse

- Suppose we want to solve $\mathbf{A}d\mathbf{Z} = d\mathbf{B}$
- $\mathbf{A}^{-1}$ is quasi-Toeplitz of form $T(\mathbf{a}^{-1}) + \mathbf{E}$, with $\mathbf{E}$ low-rank like we saw
- Key point: $\mathbf{a}^{-1}$ is **really** easy to calculate!
  - Get $a(z)$ at many z using FFT, then go from $a(z)^{-1}$ to $\mathbf{a}^{-1}$ with inverse FFT
  - Cost is only $O(T \log T)$, way cheaper than $O(T^3)$ matrix inversion
  - What can we do with just $T(\mathbf{a}^{-1})$?
  - [conceptually, $\mathbf{a}^{-1}$ is inverse for infinitely-well-anticipated shocks]

What can we do with $\mathbf{a}^{-1}$?

- Start with $(T(\mathbf{a}) + \mathbf{E})d\mathbf{Z} = d\mathbf{B}$, multiply both sides by $T(\mathbf{a}^{-1})$:

$$T(\mathbf{a}^{-1})(T(\mathbf{a}) + \mathbf{E})d\mathbf{Z} = T(\mathbf{a}^{-1})d\mathbf{B}$$

- Both $T(\mathbf{a}^{-1})T(\mathbf{a}) - \mathbf{I}$ and $T(\mathbf{a}^{-1})\mathbf{E}$ compact, well-approximated by low rank, so can be written in form

$$(\mathbf{I} + \mathbf{C})d\mathbf{Z} = d\mathbf{y}$$

- Iterative method (GMRES) **very** good at solving $(\mathbf{I} + \mathbf{C})^{-1}d\mathbf{y}$ if $\mathbf{C}$ low-rank

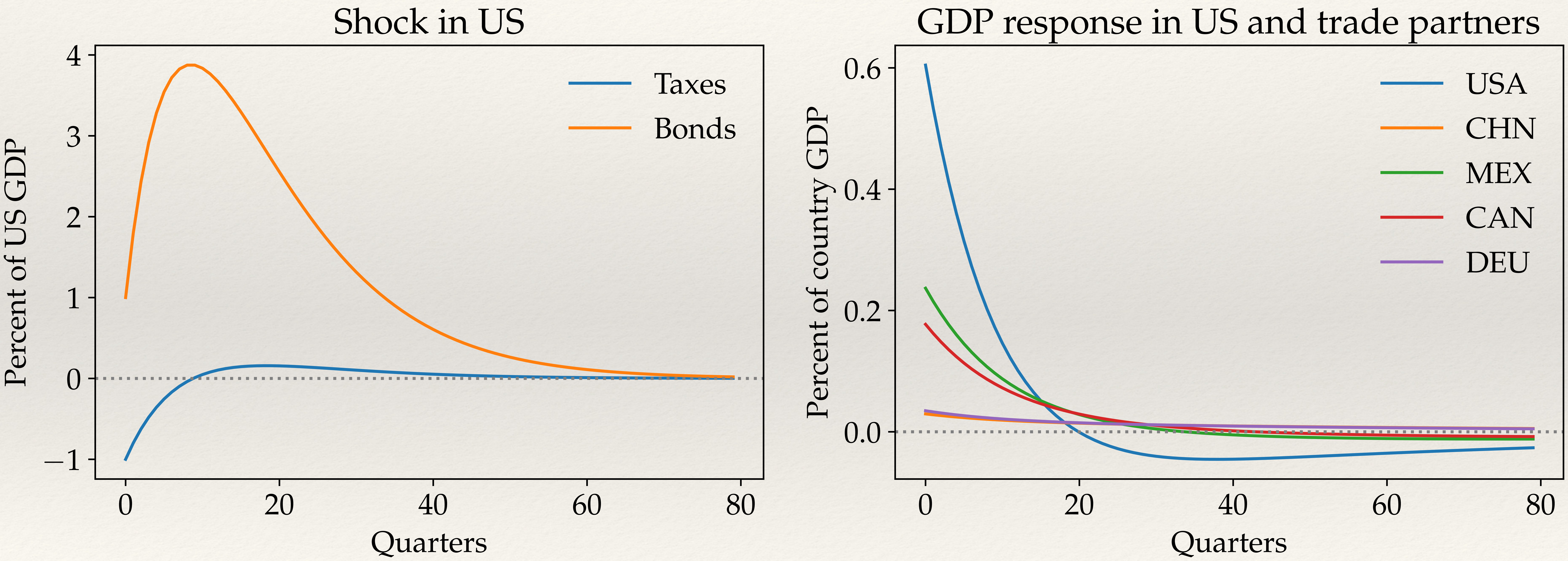
  [Multiplying by $T(\mathbf{a}^{-1})$ is called “preconditioning”.]

- Cheap, doesn’t require explicitly forming new matrices like $\mathbf{C}$

We expand this to HUGE model

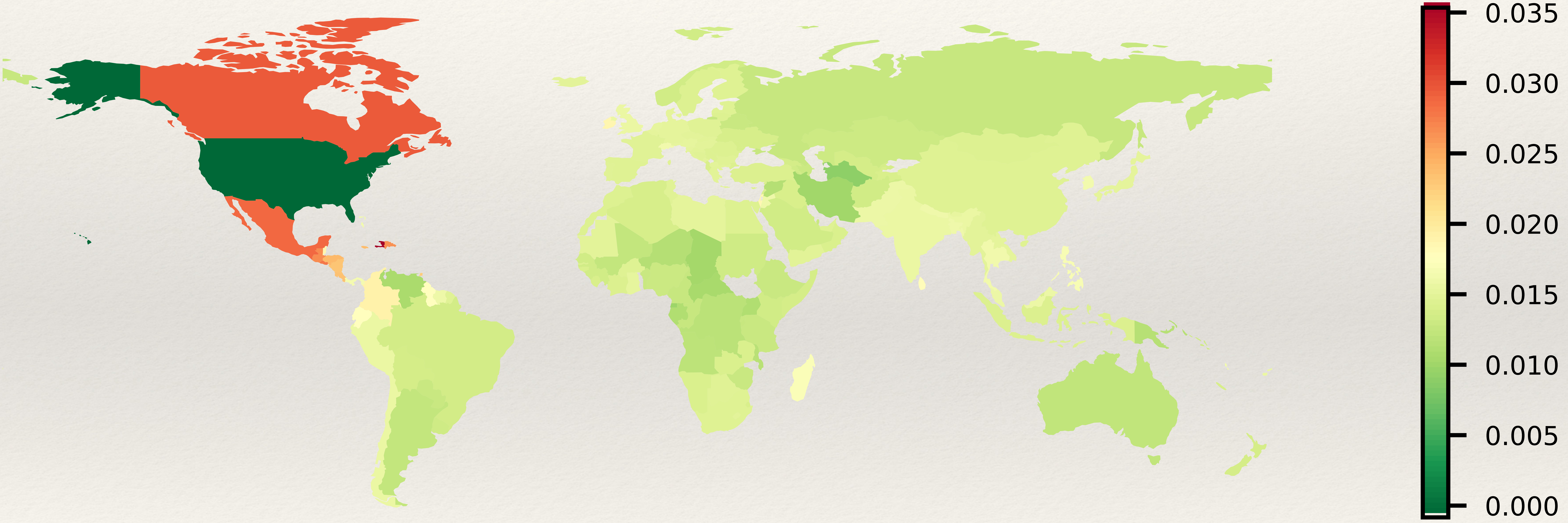
- N-country extension of IKC model, constant r in each country n
- Fiscal policy in n chooses $\{B_t^n, T_t^n\}$ consistent with budget constraint
- n spends share $\Pi_{n,n'}$ on output from others n', take from data for 177 countries
- Assume same HA model in each n, for simplicity assume all share $\mathbf{A}, \mathbf{M}$
- Solve for GDP $\{Y_{nt}\}$ in all N countries, in response to US deficit-financed tax cut, need long horizon $T = 1000$
- Usual sequence-space approach: Jacobian size $(177{,}000)^2$: can't even store!
- With iterative approach, solves in a few seconds on laptop!

Peek at solution: selected countries over time

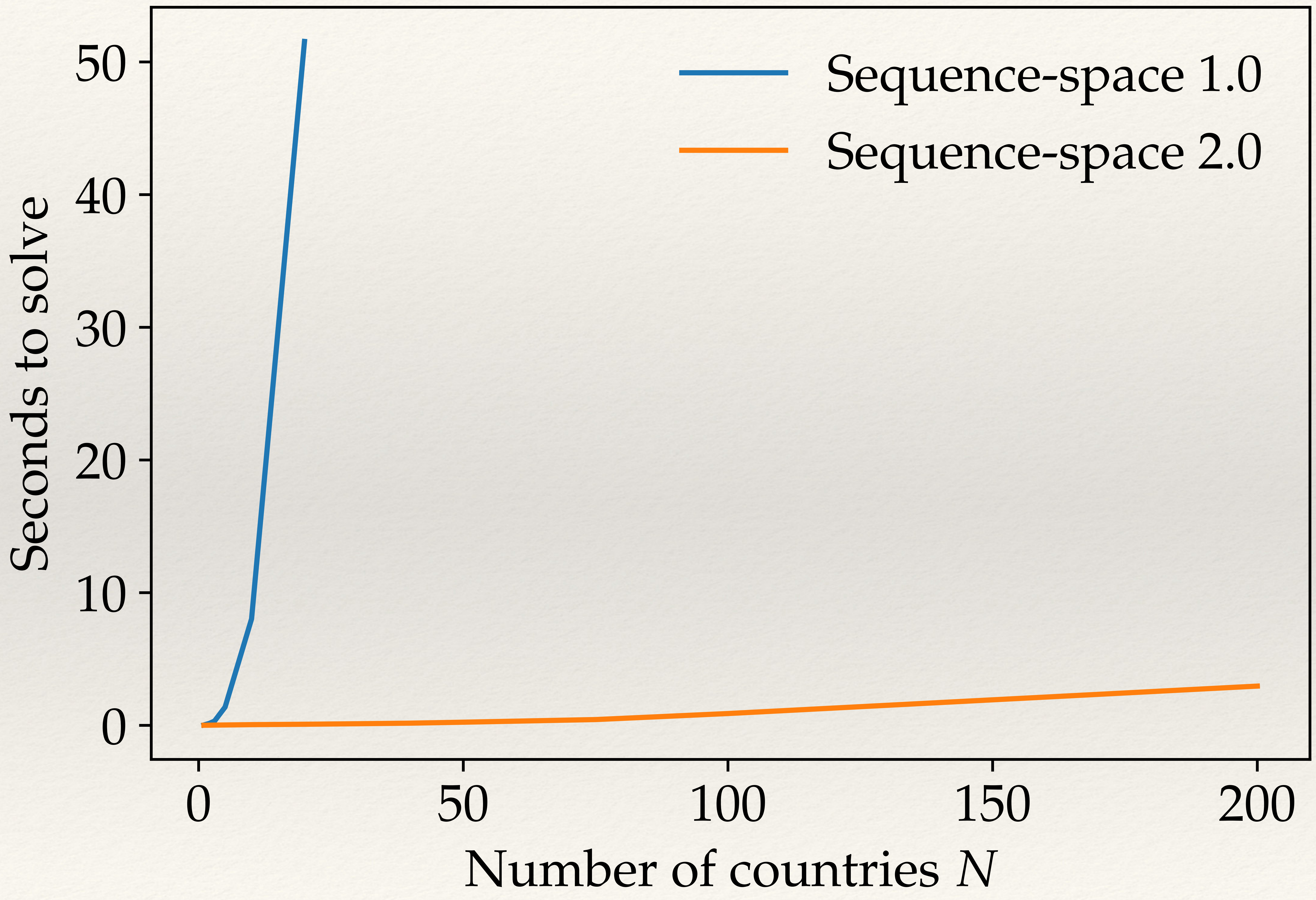


Peek at solution: on impact across countries

Peek at solution: after 20 quarters across countries

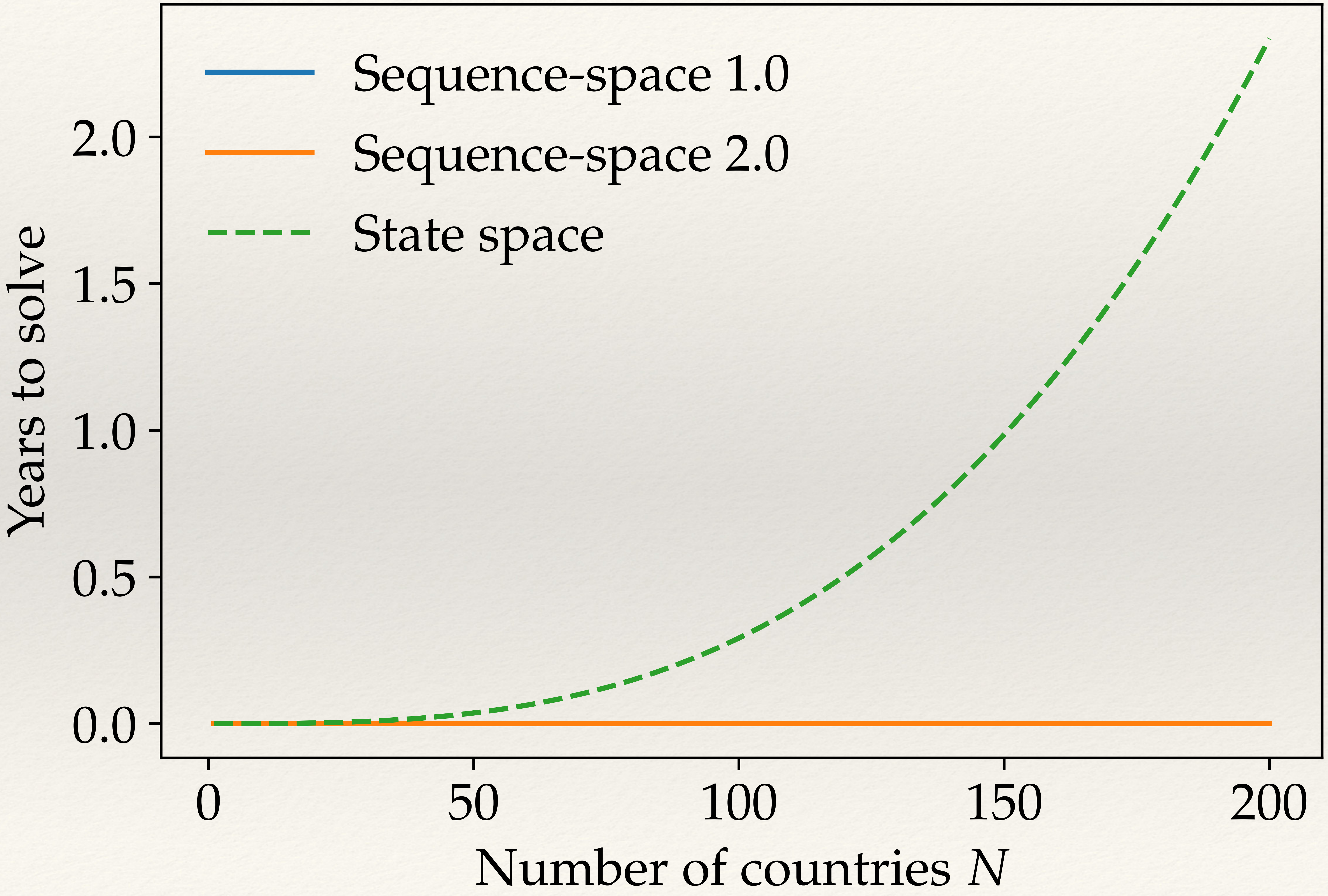


How fast is this?



How fast is this? Compare to state space

(Extrapolation using cubic cost of solving state-space system!)



Conclusion

- Quasi-Toeplitz structure of Jacobians delivers:
  - winding number test for determinacy
    - **already ready to use**!
  - faster, truncation-free computations:
    - still **in development**, bypasses major issue with sequence space
  - extremely fast iterative computations, even in huge models
    - also still **in development**, but solves 177-country HANK in 3 seconds!!