Nonlinear Structural VARs

James A. Duffy
University of Oxford

Modern Topics in Time Series Analysis
Klagenfurt am Wörthersee, September 2026

Overview

- In these two lectures, we discuss a new class of 'endogenously nonlinear' structural VARs

I. Model framework and identification (Duffy and Mavroeidis, 2026)
   - background: endogenous (v. exogenous) regime switching
   - observational equivalence and identification

II. Cointegration (Duffy and Mavroeidis, 2024)
   - reconciling common stochastic trends with nonlinear time series models
   - nonlinear extensions of the Granger–Johansen representation theorem

- A full bibliography appears at the end of these slides

Part I

Model framework and identification

Section 1

Background and motivation

Motivating questions

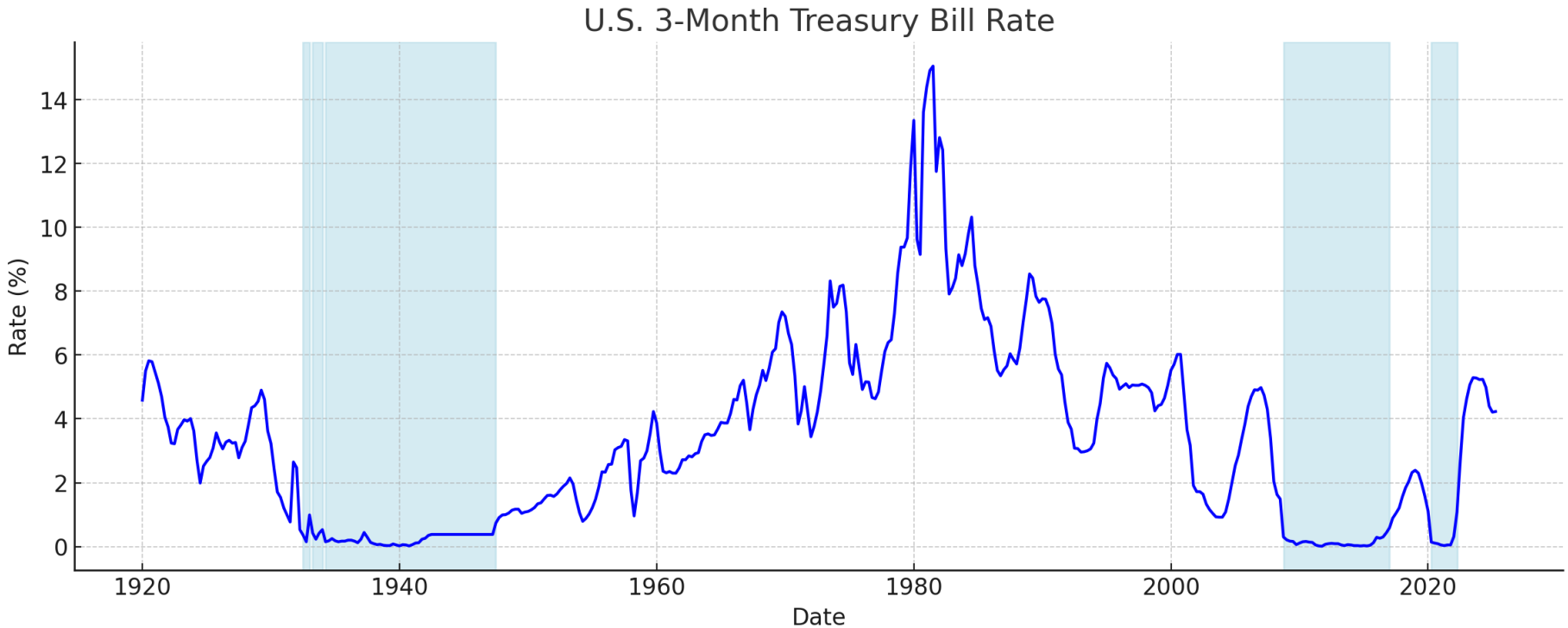
How can we account for . . .

1. Occasionally binding constraints (e.g. the 'zero lower bound')?
2. Asymmetries (e.g. in recessions v. expansions)?
3. Tipping points (e.g. financial distress, debt sustainability)?

. . . in a tractable, multivariate dynamic model?

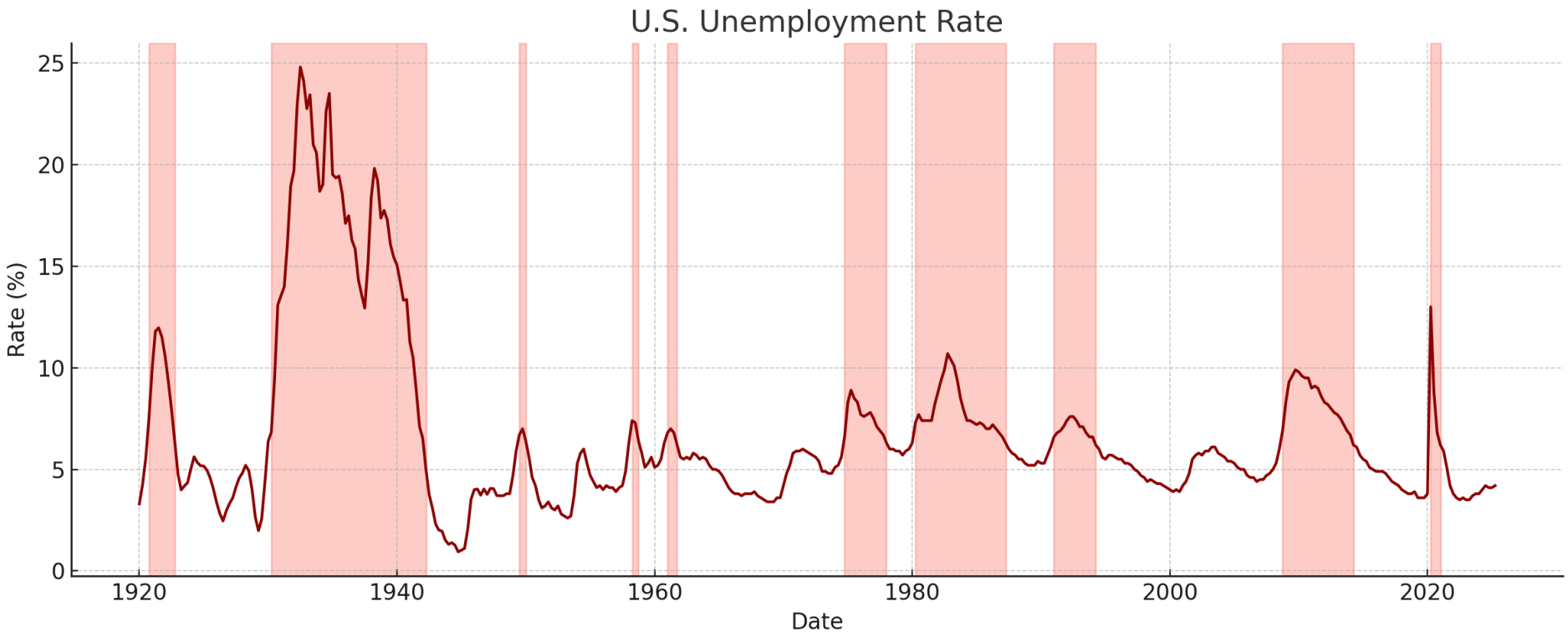
- Don't we already know how to do this in a structural VAR?

Motivation: the zero lower bound



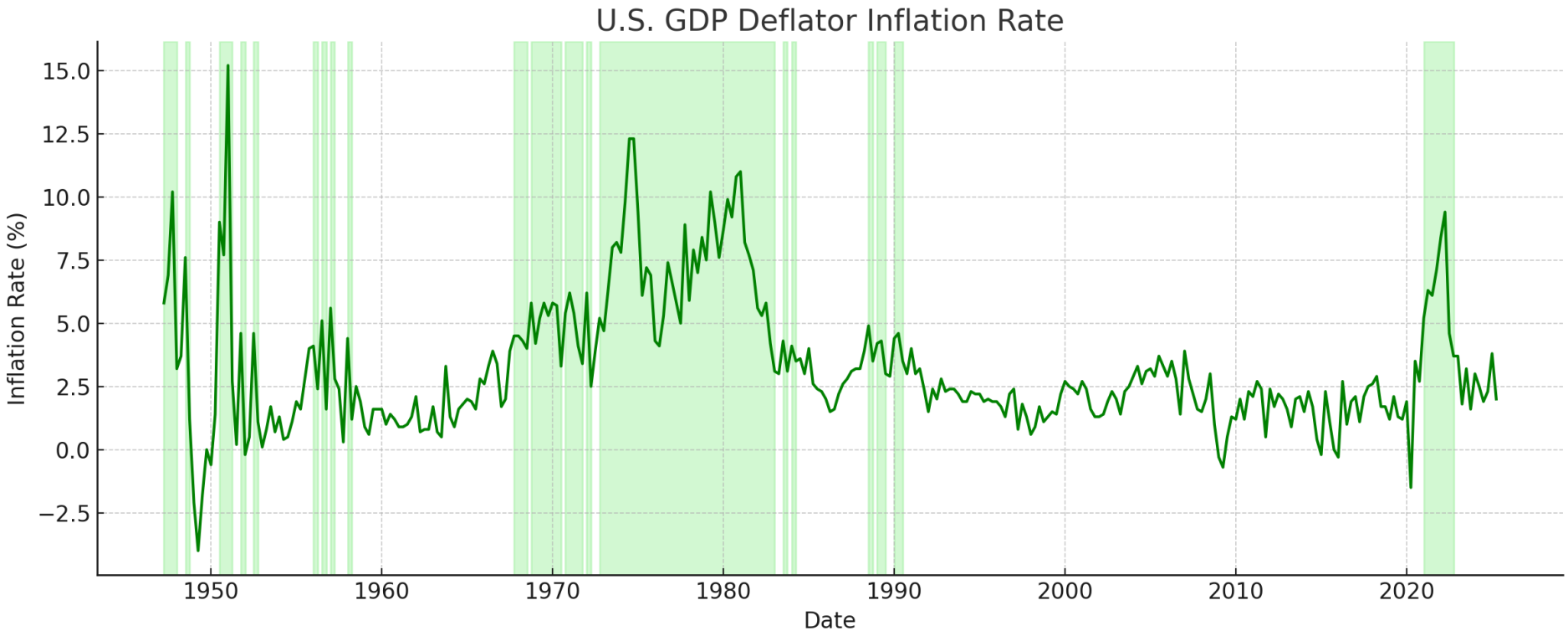
- Salience of the ZLB in recent (and some not so recent!) history
- How can we enforce non-negativity of the interest rate?

Motivation: recessions versus expansions



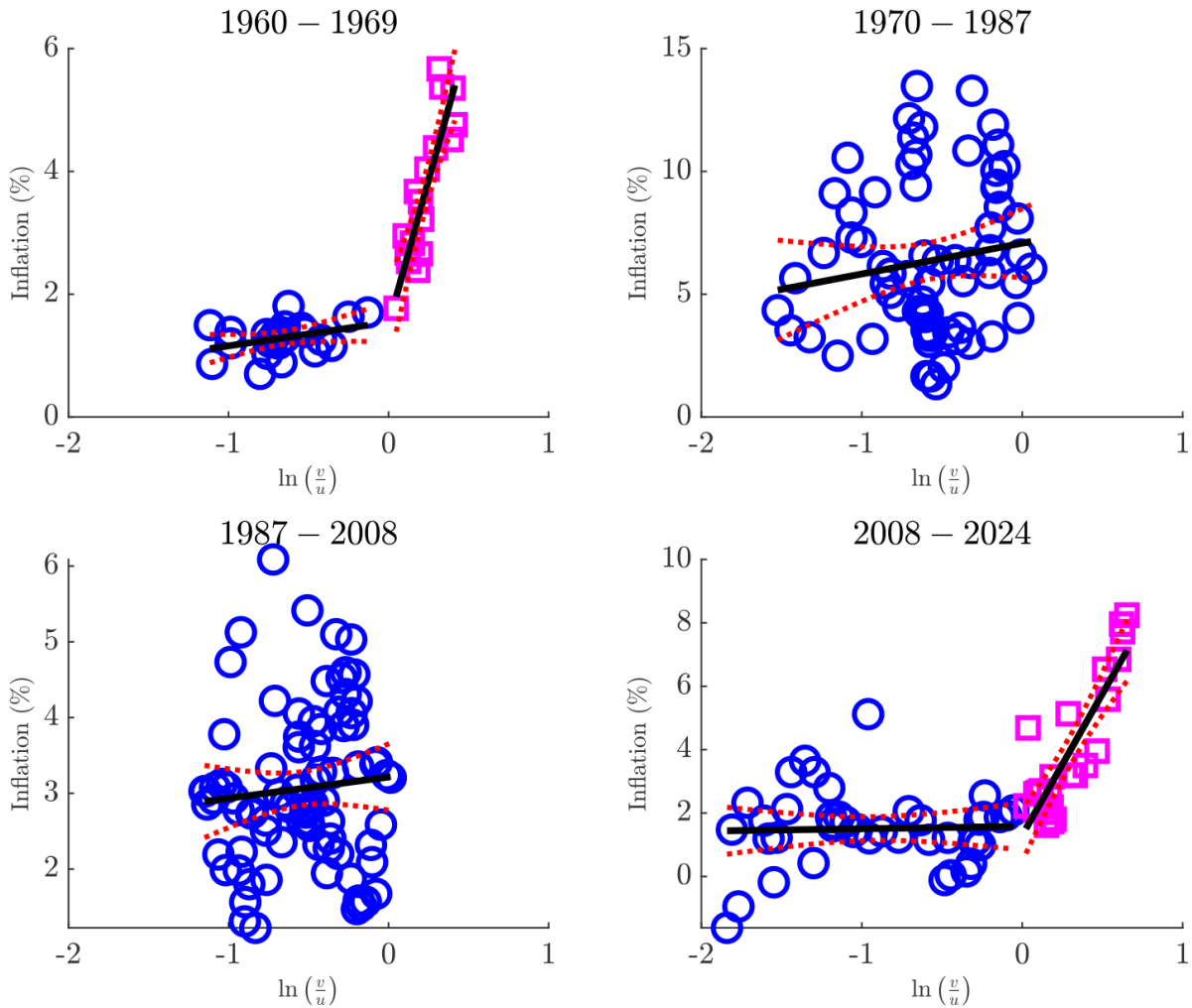
- Shaded areas correspond to unemployment rate exceeding $\tau = 6.5\%$ (Ramey and Zubairy, 2018)
- How can we allow for asymmetries (e.g. in spending multipliers) across slack ($U > \tau$) and tight ($U < \tau$) labour markets?

Motivation: a nonlinear Phillips curve?



- Ascari, Bonomolo, and Haque (2023): argue PC is steeper when (trend) inflation exceeds 4%

Motivation: a nonlinear Phillips curve?



- Benigno and Eggertsson (2023) argue for a 'Beveridge threshold'
- PC steeper when vacancies > unemployed

The existing state of the art

- Make the SVAR parameters 'regime dependent':

$$\Phi_0(s_{t-1})z_t = c(s_{t-1}) + \sum_{i=1}^{k} \Phi_i(s_{t-1})z_{t-i} + \varepsilon_t \qquad s_{t-1} \in \{1, \ldots, L\}$$

1. Threshold autoregressive models (Tong, 1990; Teräsvirta, Tjøstheim, and Granger, 2010; Hubrich and Teräsvirta, 2013)
   - e.g. with $L = 2$ discrete regimes

$$s_{t-1} = 1\{z_{1,t-1} > \tau\} + 1$$

   - or smoothly combining the two, as for some cdf F,

$$\Phi_i^{(1)}[1 - F(z_{1,t-1} - \tau)] + \Phi_i^{(2)} F(z_{1,t-1} - \tau)$$

   - in an SVAR: Auerbach and Gorodnichenko (2012); Caggiano, Castelnuovo, Colombo, and Nodari (2015)
2. Markov switching SVARs:
   - $\{s_t\}$ a Markov process either independent of all $\{\varepsilon_t\}$ (Hamilton, 1989);
   - or at least of *current and future* $\{\varepsilon_t\}$ (Hubrich and Waggoner, 2022)

The existing state of the art

$$\Phi_0(s_{t-1})z_t = \sum_{i=1}^{k} \Phi_i(s_{t-1})z_{t-i} + \varepsilon_t$$

- Unfortunately, these do not allow for genuinely *endogenous* regime switches, because the regime s_{t-1} is determined prior to ε_t
- *This matters*: e.g. for enforcing constraints and capturing contemporaneous asymmetries
- Some researchers acknowledge the limitation but worry that dealing with endogeneity is "complicated", e.g.,

  'Of course it is plausible that the variance regime switches are not only random, but endogenously determined. Allowing for that would greatly complicate the model [...] We leave this to future research.' Brunnermeier, Palia, Sastry, and Sims (2021)

- Today we discuss a new generation of SVARs with *endogenous* regime switching / nonlinear state dependence

Section 2

Endogenous regime switching

Why endogenous regime switching?

- Suppose we are interested in the effect of monetary policy shocks on unemployment and inflation
- Methodology: SVAR, with one lag ($k = 1$) for simplicity

$$\Phi_0 z_t = \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t = \Phi_1 z_{t-1} + \varepsilon_t, \qquad z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} \begin{matrix} \in \mathbb{R} \\ \in \mathbb{R}^{p-1} \end{matrix}$$

  - y_t is the policy rate, = the 'shadow rate' when negative
  - x_t collects unemployment, inflation + possibly other variables
  - $\varepsilon_t \sim_{\text{i.i.d.}} [0, I_p]$ the structural shocks, say ε_{1t} is the m.p. shock
- Causal effect of policy given by the SIRF

$$\frac{\partial z_{t+h}}{\partial \varepsilon_{1,t}} = (\Phi_0^{-1}\Phi_1)^h \Phi_0^{-1} e_1$$

  - unaffected by whether or not the ZLB binds
  - so how could we account for its possible effects . . . ?

Why exogenous regime switching fails

- Try to account for the ZLB by having two regimes, $\Phi_i^{(s_{t-1})}$ for $s_{t-1} \in \{1, 2\}$

$$\begin{aligned}\Phi_0^{(s_{t-1})} z_t &= \Phi_1^{(s_{t-1})} z_{t-1} + \varepsilon_t \\ \implies z_t &= (\Phi_0^{(s_{t-1})})^{-1} \Phi_1^{(s_{t-1})} z_{t-1} + (\Phi_0^{(s_{t-1})})^{-1} \varepsilon_t \\ &= \Psi_1^{(s_{t-1})} z_{t-1} + \Psi_0^{(s_{t-1})} \varepsilon_t\end{aligned}$$

- Can we force the shadow rate y_t to be ≤ 0 in the ZLB ($s_{t-1} = 1$) regime?

$$z_t = \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \Psi_1^{(s_{t-1})} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \Psi_0^{(s_{t-1})} \varepsilon_t$$

  - since $s_{t-1} = 1$ is determined *prior* to ε_t, this requires

$$y_t = e_1^\top z_t \leq 0 \iff e_1^\top \Psi_1^{(1)} z_{t-1} + e_1^\top \Psi_0^{(1)} \varepsilon_t \leq 0$$

  - but why should ε_t comply? Recall it is i.i.d.!
  - the coefficients for the ZLB regime would apply, even though $y_t > 0$!

Solution: endogenous regimes

- Suppose we allow the value of z_t itself to determine the current 'regime':
  - $y_t \leq 0$ is the ZLB-constrained regime ($s_t = 1$)
  - $y_t > 0$ is the unconstrained regime ($s_t = 2$)
- We allow the coefficients on y_t to differ according to their sign:
  - original linear model

$$\Phi_0 z_t = \begin{bmatrix} \phi_0 & \Phi_0^x \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \phi_0 y_t + \Phi_0^x x_t$$

  - becomes

$$\Phi_0(z_t) z_t = \begin{cases} \Phi_0^{(1)} z_t & \text{if } y_t \leq 0 \\ \Phi_0^{(2)} z_t & \text{if } y_t > 0 \end{cases} \qquad \begin{aligned} \Phi_0^{(1)} &= [\phi_0^- \quad \Phi^x] \\ \Phi_0^{(2)} &= [\phi_0^+ \quad \Phi^x] \end{aligned}$$

  - with regime indicator determined by y_t, and thus by ε_t:

$$s_t = 1\{y_t > 0\} + 1$$

- Yields the *endogenous* regime-switching model

$$\Phi_0^{(s_t)} z_t = \Phi_1^{(s_{t-1})} z_{t-1} + \varepsilon_t$$

The CKSVAR model

- Mavroeidis (2021): the censored and kinked structural VAR:

$$\phi_0^+ y_t^+ + \phi_0^- y_t^- + \Phi_0^x x_t = c + \sum_{i=1}^{k} [\phi_i^+ y_{t-i}^+ + \phi_i^- y_{t-i}^- + \Phi_i^x x_{t-i}] + \varepsilon_t$$

  (cf. Aruoba, Mlikota, Schorfheide, and Villalvazo, 2022)
- Interpretation? If y_t is the 'stance of monetary policy' ...
  - y_t^+ is the actual (observed) policy rate: constrained to be nonnegative
  - y_t^- is the 'shadow rate': a measure of unconventional policy
- Two endogenously switching regimes on the l.h.s.
  - is continuous in $z_t = (y_t, x_t^\top)^\top$
  - invertible under conditions given in Gouriéroux, Laffont, and Monfort (1980)
  - necessary and sufficient that

$$\Phi_0^+ := [\phi_0^+, \Phi_0^x] \qquad \Phi_0^- := [\phi_0^-, \Phi_0^x]$$

  each be invertible, with determinants having the same sign

The CKSVAR model: asymmetries

$$\Phi_0(z_t)z_t = c + \sum_{i=1}^{k} \Phi_i(z_{t-i})z_{t-i} + \varepsilon_t$$

- Impact multipliers are given by

$$\frac{\partial z_t}{\partial \varepsilon_t} = \Phi_0(z_t)^{-1} = \begin{cases} (\Phi_0^{(1)})^{-1} & \text{if } y_t < 0 \\ (\Phi_0^{(2)})^{-1} & \text{if } y_t > 0 \end{cases}$$

- So if y_t is at the threshold (here, $y_t = 0$), responses to ε_t will be *asymmetric*
- Not possible if regime predetermined, because e.g. $\Phi_0(z_{t-1})^{-1}$ would not depend on ε_t

A general nonlinear framework

- Additively time-separable nonlinear SVAR:

$$f_0(z_t) = c + \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

  - f_0 is a homeomorphism
  - $\{f_i\}_{i=0}^{k}$ are continuous
- Accommodates the CKSVAR:

$$\begin{aligned} f_i(z) = f_i(y, x) &= \phi_i^+ [y]_+ + \phi_i^- [y]_- + \Phi_i^x x \\ &= [\Phi_i^+ 1\{y \geq 0\} + \Phi_i^- 1\{y < 0\}] z \end{aligned}$$

- Major novelty is the nonlinearity on the l.h.s.
  - nonlinearity in the endogenous variables; an 'endogenously nonlinear' SVAR
  - no longer a regression model
  - requires *independence* of ε_t from $\{z_{t-i}\}_{i \geq 1}$, not mere orthogonality

Multiple endogenous regimes

$$f_0(z_t) = c + \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

- With f_i piecewise affine (and continuous):

$$f_i(z) = \sum_{\ell=1}^{L} 1\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z)$$

  - $\{\mathscr{Z}^{(\ell)}\}_{\ell=1}^{L}$ are convex and partition $\mathbb{R}^p$; demarcate L regimes
  - invertibility of f_0: again, see Gouriéroux et al. (1980)
- Propose to use a convolution to smoothly interpolate regimes

$$f_{i,K}(z) := \int_{\mathbb{R}^p} f_i(z+e)K(e)\,\mathrm{d}e$$

  - $f_{0,K}$ is (under our conditions) invertible if unsmoothed f_0 is
  - not the case if $1\{z \in \mathscr{Z}^{(\ell)}\}$ simply replaced by some cdf-like function

- Implemented by **EndoRSE** (Endogenous Regime-Switching model Estimation) Julia programme (Bonomolo, Guarnieri, Kabel & Mavroeidis)

Section 3

Observational equivalence and identification

Identification: in a linear SVAR

- Consider the linear SVAR

$$\Phi_0 z_t = c + \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t, \qquad \varepsilon_t \sim [0, I_p], \qquad \mathbb{E}\varepsilon_t z_s^\top = 0,\ s < t$$

  - a dynamic SEM: $\{z_{t-i}\}_{i\geq 1}$ are the exogenous (predetermined) variables
- Identification can be inferred from the reduced form

$$z_t = \Phi_0^{-1} c + \sum_{i=1}^{k} \Phi_0^{-1}\Phi_i z_{t-i} + u_t \qquad u_t \sim [0, (\Phi_0^\top \Phi_0)^{-1}]$$

  - $(\tilde{c}, \{\tilde{\Phi}_i\})$ indistinguishable from $(c, \{\Phi_i\})$ if yield the same reduced form
  - true iff there exists an orthogonal $Q \in \mathbb{O}(p)$ such that

$$\tilde{c} = Qc \qquad \tilde{\Phi}_i = Q\Phi_i,\ \forall i$$

- Fundamental identification problem in macroeconomics: we (always) need further restrictions to pin down Q

Identification in the nonlinear SVAR

$$f_0(z_t) = \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t =: \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t$$

- Because f_0 now possibly nonlinear, need to strengthen some conditions

1. ε_t is *independent* of $\boldsymbol{z}_{t-1}$
   - i.e. not merely uncorrelated
   - $\varepsilon_t \sim_{\text{i.i.d.}} [0, I_p]$, with continuous density ϱ, having a 'regular' local maximum
2. $\boldsymbol{f}_1 : \mathbb{R}^{kp} \to \mathbb{R}^p$ must depend 'sufficiently' on $\boldsymbol{z}_{t-1}$:
   - $\boldsymbol{z} \mapsto \boldsymbol{f}_1(\boldsymbol{z})$ is surjective (i.e. onto)
   - $D\boldsymbol{f}_1(\boldsymbol{z})$ has rank p a.e.
3. f_0, f_0^{-1} and $\boldsymbol{f}_1$ locally Lipschitz [automatic in linear case]

- Other key (implicit) assumption is separability in endogenous (z_t) and predetermined ($\boldsymbol{z}_{t-1}$) variables (of course true in the linear SVAR!)
- We make weaker assumptions on the 'alternative' parameters $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$

Characterising observational equivalence

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad\qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- What do we mean by observational equivalence here?
  - $(f_0, \boldsymbol{f}_1, \varrho)$ determine the conditional density $z_t \mid \boldsymbol{z}_{t-1}$

$$\varphi_{z_t|\boldsymbol{z}_{t-1}}(\xi \mid \boldsymbol{\xi}_{-1}) = \varrho[f_0(\xi) - \boldsymbol{f}_1(\boldsymbol{\xi}_{-1})] \cdot |\det Df_0(\xi)|$$

  - $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1, \tilde{\varrho})$ is obs. equivalent to $(f_0, \boldsymbol{f}_1, \varrho)$ if it yields the same density

- Q. For a given $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$: when does there exist a $\tilde{\varrho}$ such that $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1, \tilde{\varrho})$ is observationally equivalent to $(f_0, \boldsymbol{f}_1, \varrho)$?

- Problem relates to the microeconometrics literature on nonlinear SEMs; particularly Matzkin (2008), Berry and Haile (2018)

Identification: main result

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- For a given $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$: there exists a $\tilde{\varrho}$ such that $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1, \tilde{\varrho})$ is observationally equivalent to $(f_0, \boldsymbol{f}_1, \varrho)$, iff $\exists Q \in \mathbb{O}(p)$ such that

$$\tilde{f}_0(z) = Q f_0(z), \ \forall z \in \mathbb{R}^p \qquad \tilde{\boldsymbol{f}}_1(\boldsymbol{z}) = Q \boldsymbol{f}_1(\boldsymbol{z}), \ \forall \boldsymbol{z} \in \mathbb{R}^{kp}$$

- *Exactly* as in a linear SVAR:
  - parameters $(f_0, \boldsymbol{f}_1)$ are identified up to an unknown $Q \in \mathbb{O}(p)$
  - 'identification problem' is the same: greater flexibility comes at no cost!
  - only $p(p-1)/2$ restrictions needed to achieve exact identification
  - a 'nonparametric' result: no functional form assumptions

Why independence?

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- Why do we need independence of ε_t from $\boldsymbol{z}_{t-1}$?
- Allows the density $z_t \mid \boldsymbol{z}_{t-1}$ to be determined by the model
- Allows us to exclude parameters $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$ for which

$$\tilde{\varepsilon}_t = \tilde{f}_0(z_t) - \tilde{\boldsymbol{f}}_1(\boldsymbol{z}_{t-1}) \not\perp \boldsymbol{z}_{t-1}$$

  - thus yielding key identifying restrictions

Why 'sufficient dependence'?

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- Why do we need 'sufficient dependence' of $\boldsymbol{f}_1$ on $\boldsymbol{z}_{t-1}$?
- Suppose $\boldsymbol{f}_1(\boldsymbol{z}) = 0$, so that

$$f_0(z_t) = \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p]$$

- Now consider *any* (invertible) $\tilde{f}_0$:

$$\tilde{\varepsilon}_t := \tilde{f}_0(z_t) = \tilde{f}_0[f_0^{-1}(\varepsilon_t)]$$

  - must also be i.i.d., and independent of $\boldsymbol{z}_{t-1}$
  - may be re-normalised to have mean zero and variance I_p
- Essentially *nothing* about f_0 would be identified!

Outline of proof

$$\varrho[f_0(z) - \boldsymbol{f}_1(\boldsymbol{z}_{-1})] \cdot \det Df_0(z) = \tilde{\varrho}[\tilde{f}_0(z) - \tilde{\boldsymbol{f}}_1(\boldsymbol{z}_{-1})] \cdot \det D\tilde{f}_0(z)$$

1. Show $\det Df_0(z) = \mathrm{e}^C \cdot \det D\tilde{f}_0(z)$ $\quad \Longleftarrow$ separability of z_t, $\boldsymbol{z}_{t-1}$
2. Relate $\tilde{\boldsymbol{f}}_1$ to $\boldsymbol{f}_1$ via $\tilde{m}_0 = \tilde{f}_0 \circ f_0^{-1}$ $\quad \Longleftarrow$ local max at $\varrho(\varepsilon^*)$

$$\tilde{\boldsymbol{f}}_1(\boldsymbol{z}_{-1}) = \tilde{m}_0[\varepsilon^* + \boldsymbol{f}_1(\boldsymbol{z}_{-1})] - \tilde{\varepsilon}^*$$

3. Show $\tilde{m}_0$ is affine $\quad \Longleftarrow$ local max + rigidity theorem

$$[D\tilde{m}_0(w)^\top]^{-1}[D^2\varrho(\varepsilon^*)][D\tilde{m}_0(w)]^{-1} = S$$

4. Deduce

$$\tilde{\varepsilon}_t = \tilde{f}_0(z_t) - \tilde{\boldsymbol{f}}_1(\boldsymbol{z}_{t-1}) = q_0 + Q[f_0(z_t) - \boldsymbol{f}_1(\boldsymbol{z}_{t-1})] = q_0 + Q\varepsilon_t$$

whence $q_0 = 0$ and $QQ^\top = I_p$

Section 4

Discussion: implications and extensions

Implications for exact identification

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- Fix an (arbitrary) $z^* \in \mathbb{R}^p$ and (uniquely, wlog) decompose

$$Df_0(z^*) = Q^\top L \qquad Q \in \mathbb{O}(p),\ L \text{ lower triangular, pos. diag}$$

- By our result: L is identified, Q is not
- We may reparametrise the system as

$$g_0(z_t) = \boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + Q\varepsilon_t$$

  - where $(g_0, \boldsymbol{g}_1) := (Qf_0, Q\boldsymbol{f}_1)$ *is* identified
  - $p(p-1)/2$ external restrictions are needed to pin down (all of) Q

- Existing approaches to identification in a linear SVAR (timing, sign restrictions, instruments, etc.) apply directly to this model

Identification of impulse responses

$$g_0(z_t) = \boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + Q\varepsilon_t =: \boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + u_t$$

- We have identified $(g_0, \boldsymbol{g}_1)$, and thus also the 'nonstructural' shocks $\{u_t\}$
- Recursive substitution of

$$z_t = g_0^{-1}[\boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + u_t] =: m_0[u_t, \boldsymbol{z}_{t-1}; g_0, \boldsymbol{g}_1]$$

yields the (identified) mapping

$$z_{t+h} = m_h[u_t, \{u_s\}_{s=t+1}^{t+h}, \boldsymbol{z}_{t-1}; g_0, \boldsymbol{g}_1]$$

- Structural impulse responses (e.g. Koop, Pesaran, and Potter, 1996):
  - compare baseline and counterfactual paths for $h \mapsto z_{t+h}$, given

$$\varepsilon_{1t} = e \qquad \text{v.} \qquad \varepsilon_{1t} = e + \delta$$

  - averaging out other current $\varepsilon_{-1,t}$ and future $\{\varepsilon_s\}_{s=t+1}^{t+h}$ structural shocks
  - here conditional on $\boldsymbol{z}_{t-1}$; but possibly also averaging over $\boldsymbol{z}_{t-1}$

Identification of impulse responses

$$\Psi_h(\delta; e, \zeta) := \mathbb{E}[z_{t+h} \mid \varepsilon_{1t} = e + \delta, \boldsymbol{z}_{t-1} = \zeta] - \mathbb{E}[z_{t+h} \mid \varepsilon_{1t} = e, \boldsymbol{z}_{t-1} = \zeta]$$

- Conditional on $\varepsilon_{1t} = \bar{\varepsilon}_1$, we have to integrate out from

$$z_{t+h} = m_h[u_t, \{u_s\}_{s=t+1}^{t+h}, \zeta; g_0, \boldsymbol{g}_1]$$

  - future $\{u_s\}_{s=t+1}^{t+h}$ $\implies$ ok, because independent of ε_{1t}
  - current u_t *conditional* on $\varepsilon_{1t} = \bar{\varepsilon}_1$ $\implies$ a problem?

- Suppose we can identify only the first column q_1 of Q in

$$u_t = Q\varepsilon_t = q_1\varepsilon_{1t} + Q_{-1}\varepsilon_{-1,t}$$

  - then we have identified $\varepsilon_{1t} = q_1^\top u_t$
  - and thus the conditional distribution

$$u_t \mid (\varepsilon_{1t} = \bar{\varepsilon}_1) =_d u_t \mid (q_1^\top u_t = \bar{\varepsilon}_1)$$

  - and thus Ψ_h is identified

Example: identifying q_1 only

$$g_0(z_t) = \boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + Q\varepsilon_t = \boldsymbol{g}_1(\boldsymbol{z}_{t-1}) + q_1\varepsilon_{1t} + Q_{-1}\varepsilon_{-1,t}$$

- So to identify the SIRF w.r.t. ε_{1t}: we need only q_1

1. Short-run restrictions:
   - suppose in state $z_t = z^*$, ε_{1t} is the only shock that affects z_{1t} on impact
   - because $Dg_0(z^*)$ is lower triangular, it follows that $q_1 = e_1$
   - approach extends more generally to 'block recursive' identification schemes
2. External instruments
   - suppose w_t is a scalar instrument, correlated *only* with ε_{1t}
   - since u_t is identified, q_1 is identified from

$$\mathbb{E}u_t w_t = Q\mathbb{E}\varepsilon_t w_t = Qe_1\alpha = q_1\alpha$$

Implications for parametric models

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}) + \varepsilon_t \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p], \text{ density } \varrho$$

- In practice, model is likely to be formulated parametrically, e.g.
  - $\boldsymbol{f}_1(\boldsymbol{z}_{t-1}) = c + \sum_{i=1}^{k} f_i(z_{t-i})$
  - with each f_i piecewise affine (and continuous):

$$f_i(z) = \sum_{\ell=1}^{L} 1\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z)$$

  - form of ϱ may be assumed, e.g. Gaussian
- Our result implies:
  - these parametric models are also identified up to a $Q \in \mathbb{O}(p)$
  - so parametric restrictions simplify estimation, but *not* identification
- We provide verifiable conditions for identification in piecewise affine models

Extensions

$$f_0(z_t) = \boldsymbol{f}_1(\boldsymbol{z}_{t-1}^{(1)}, \boldsymbol{z}_{t-1}^{(2)}, v_{t-1}) + \sigma(\boldsymbol{z}_{t-1}^{(2)}, v_{t-1})\varepsilon_t, \qquad \varepsilon_t \sim_{\text{i.i.d.}} [0, I_p]$$

- Augmented model allows for:
  1. ARCH-type heteroskedasticity: diagonal matrix $\sigma(\cdot)$ depends on predetermined variables
  2. Dependence on exogenous processes: $\{v_t\}$ (e.g. 'time')
  3. Discontinuities: require continuity *only* in $\boldsymbol{z}^{(1)} \mapsto \boldsymbol{f}_1(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, v)$
- Can effectively apply our previous theorem 'conditional' on each $(\boldsymbol{z}^{(2)}, v)$

- $(f_0, \boldsymbol{f}_1, \sigma)$ is identified up to a $Q \in \mathbb{O}(p)$ such that

  $$Q\sigma^2(\boldsymbol{z}^{(2)}, v)Q^\top$$

  is a diagonal matrix for 'a.e.' $(\boldsymbol{z}^{(2)}, v) \in \mathbb{R}^{d_{(2)}} \times \mathcal{V}$

- Yields 'identification by heteroskedasticity' type restrictions (if desired)

Identification and consistent estimation

- Really we have asked: what does the (a.e.) equality

$$\varrho[f_0(\xi) - \boldsymbol{f}_1(\boldsymbol{\xi}_{-1})] \cdot |\det Df_0(\xi)| = \tilde{\varrho}[\tilde{f}_0(\xi) - \tilde{\boldsymbol{f}}_1(\boldsymbol{\xi}_{-1})] \cdot |\det D\tilde{f}_0(\xi)|$$

  imply about the relationship between $(f_0, \boldsymbol{f}_1)$ and $(\tilde{f}_0, \tilde{\boldsymbol{f}}_1)$?
- Entails $\varphi_{z_t|\boldsymbol{z}_{t-\mathbf{1}}}$ is the same under these two parametrisations
  - if (a.e.) equality fails, conditional densities would be distinct
  - but could we detect these differences w.p.a.1?
- What is the connection between identification and consistent (nonparametric) estimation?
  - immediate in the i.i.d. setting: consistent estimation $\iff$ identified
  - carries over if e.g. $\{z_t\}$ is geometrically ergodic (for conditions, see Duffy, Mavroeidis, and Wycherley, 2023)
  - might not hold if e.g. $\{z_t\}$ has stochastic trends

Section 5

Application: a nonlinear Phillips curve?

A nonlinear Phillips curve?

- Benigno and Eggertsson (2023): argue for a nonlinear 'Phillips curve'

$$\pi_t - \pi = \begin{cases} \kappa \log \theta_t + \eta_{\pi,t}, & \text{if } \theta_t \leq 1 \text{ ('normal')} \\ \kappa^{\text{tight}} \log \theta_t + \eta_{\pi,t}, & \text{if } \theta_t > 1 \text{ ('tight')} \end{cases}$$

  - π_t = core CPI inflation; $\theta_t = v_t/u_t$ is the ratio of job vacancies to unemployed workers
  - use this nonlinearity to explain the post-pandemic inflation surge
- Beaudry, Hou, and Portier (2025): caution this evidence is not robust to: functional form assumptions; alternative treatments of expectations
- Evidence in these papers relies on single-equation analyses
  - potentially contingent on identifying assumptions
  - can we do better by allowing simultaneous determination of π_t and $\log \theta_t$?

A nonlinear Phillips curve?

- We consider an endogenous regime-switching SVAR for $z_t = (\log \theta_t, \pi_t)^\top$,

$$\Phi_0^{(\ell_t)} z_t = c + \sum_{i=1}^{k} \Phi_i^{(\ell_{t-i})} z_{t-i} + \varepsilon_t$$

  - regime determined by the labour market tightness $z_{1t} = \log \theta_t$:

$$\ell_t := \begin{cases} 1, & \text{if } z_{1t} \leq 0 \text{ ('normal')} \\ 2, & \text{if } z_{1t} > 0 \text{ ('tight')} \end{cases}$$

  - regime-dependent coefficient matrices: for $\ell \in \{1, 2\}$,

$$\Phi_i^{(\ell)} = \begin{bmatrix} \Phi_{i,11}^{(\ell)} & \Phi_{i,12}^{(\ell)} \\ \Phi_{i,21}^{(\ell)} & \Phi_{i,22}^{(\ell)} \end{bmatrix} = \begin{bmatrix} \Phi_{i,11}^{(\ell)} & \Phi_{i,12} \\ \Phi_{i,21}^{(\ell)} & \Phi_{i,22} \end{bmatrix}$$

- Identification via a short-run restriction?
  - suppose π_t does not contemporaneously affect $\log \theta_t$ ($\Phi_{0,12} = 0$)
  - one restriction, so model is exactly identified
  - corresponds to the identifying assumption of Benigno and Eggertsson (2023)

Nonlinearity: identification-agnostic tests

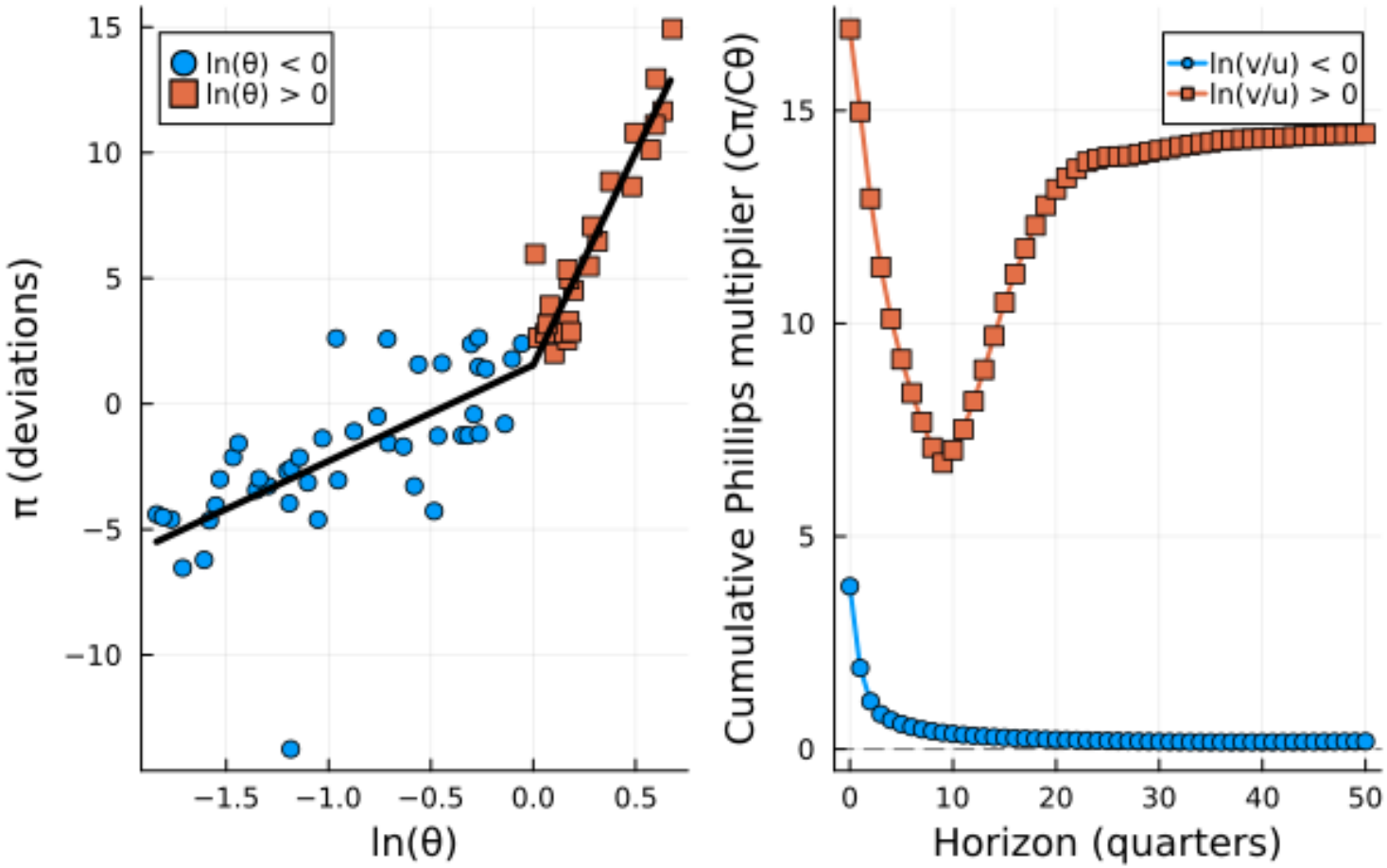
$$\Phi_0^{(\ell_t)} z_t = c + \sum_{i=1}^{k} \Phi_i^{(\ell_{t-i})} z_{t-i} + \varepsilon_t, \qquad \Phi_i^{(\ell)} = \begin{bmatrix} \Phi_{i,11}^{(\ell)} & \Phi_{i,12} \\ \Phi_{i,21}^{(\ell)} & \Phi_{i,22} \end{bmatrix}$$

- Linearity tests *agnostic* to identifying assumptions
  - nonlinearity remains in the system irrespective of the Q that rotates it
  - not necessarily true of an individual equation in the system

	LR Statistic [p-value]	
Null Hypothesis	1960–2024	2008–2024
No *endogenous* switching: $\Phi_0^{(1)} = \Phi_0^{(2)}$	21.7 [0.0]	38.6 [0.0]
Linear SVAR: $\Phi_i^{(1)} = \Phi_i^{(2)},\ \forall i \in \{0,1,2\}$	38.0 [0.0]	51.2 [0.0]

- Linearity strongly rejected, with more evidence post 2008

Estimates of the identified Phillips Curve



- Regime-dependent PC slopes: $\beta^{(\ell)} = -\Phi^{(\ell)}_{0,21}/\Phi_{0,22}$
  - for 2008–2024: $\hat{\beta}^{(1)} = 3.8$ ('normal'); $\hat{\beta}^{(2)} = 16.9$ ('tight').
- Cumulative Phillips multipliers:
  - cumulative IRFs of inflation / tightness, to a one-unit tightness shock
  - starting 2009Q3 ($\log\theta = -1.84$) and 2022Q2 ($\log\theta = +0.68$)

Part II

Stochastic trends and cointegration

Implied time series properties

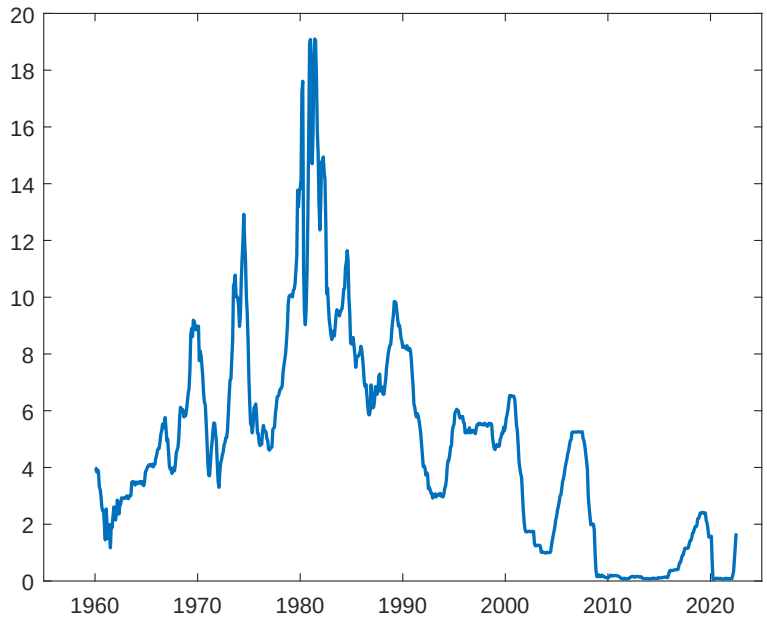
$$f_0(z_t) = c + \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

- When are $\{f_i\}_{i=0}^{k}$ such that
  1. $\{z_t\}$ is stationary and ergodic?
  2. $\{z_t\}$ has q (common) stochastic trends?
- Why is this important?
  - model adequacy: can the model match stationary and/or stochastically trending series?
  - needed to derive asymptotics of estimators and test statistics
- For piecewise affine (incl. smoothed) models, stationarity follows from stability of the deterministic, homogeneous system (Duffy et al., 2023):

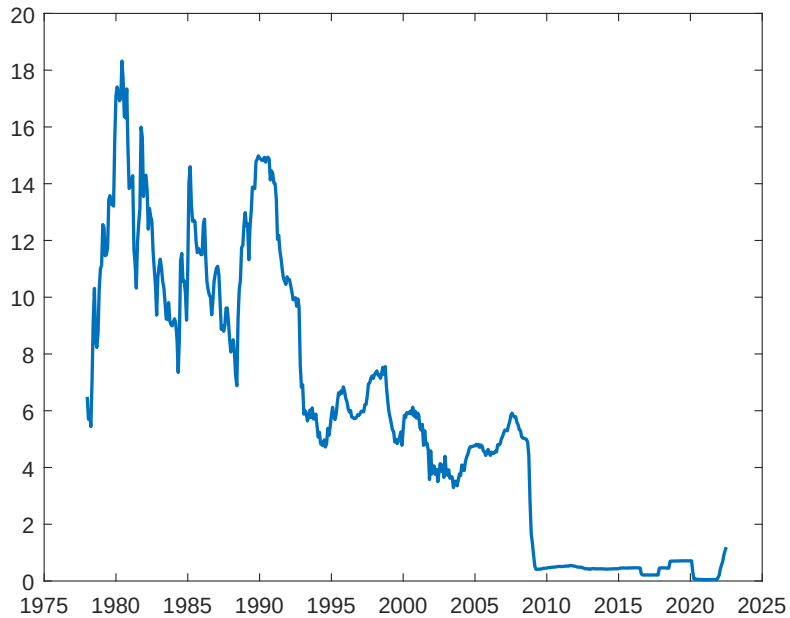
$$f_0(\hat{z}_t) = \sum_{i=1}^{k} f_i(\hat{z}_{t-i})$$

- What about (common) stochastic trends?

Common stochastic trends



US: Fed Funds rate



UK: overnight interbank rate

- Stationarity may not be plausible for many of the series that we want to model, due to evident random wandering / stochastic trends
- Want to model these series directly, without e.g. 'differencing to stationarity' $\implies$ induces misspecification

Common stochastic trends

Open problem

How can a (nonlinear) VAR generate *nonlinear* cointegration?

- Huge literatures on 'nonlinear cointegration': but they don't solve this problem! (noted by Tjøstheim, 2020)
- What do we mean by 'nonlinear cointegration'?
  - processes share common *nonlinear* stochastic trends
  - cointegrating relations may differ across regimes defined by (levels of) the integrated processes (not by stationary variables)
- We answer this question by extending the Granger–Johansen representation theorem (GJRT) to our class of nonlinear SVARs
  - here we consider the general, additively time-separable SVAR
  - Duffy, Mavroeidis, and Wycherley (2025): results specific to the CKSVAR

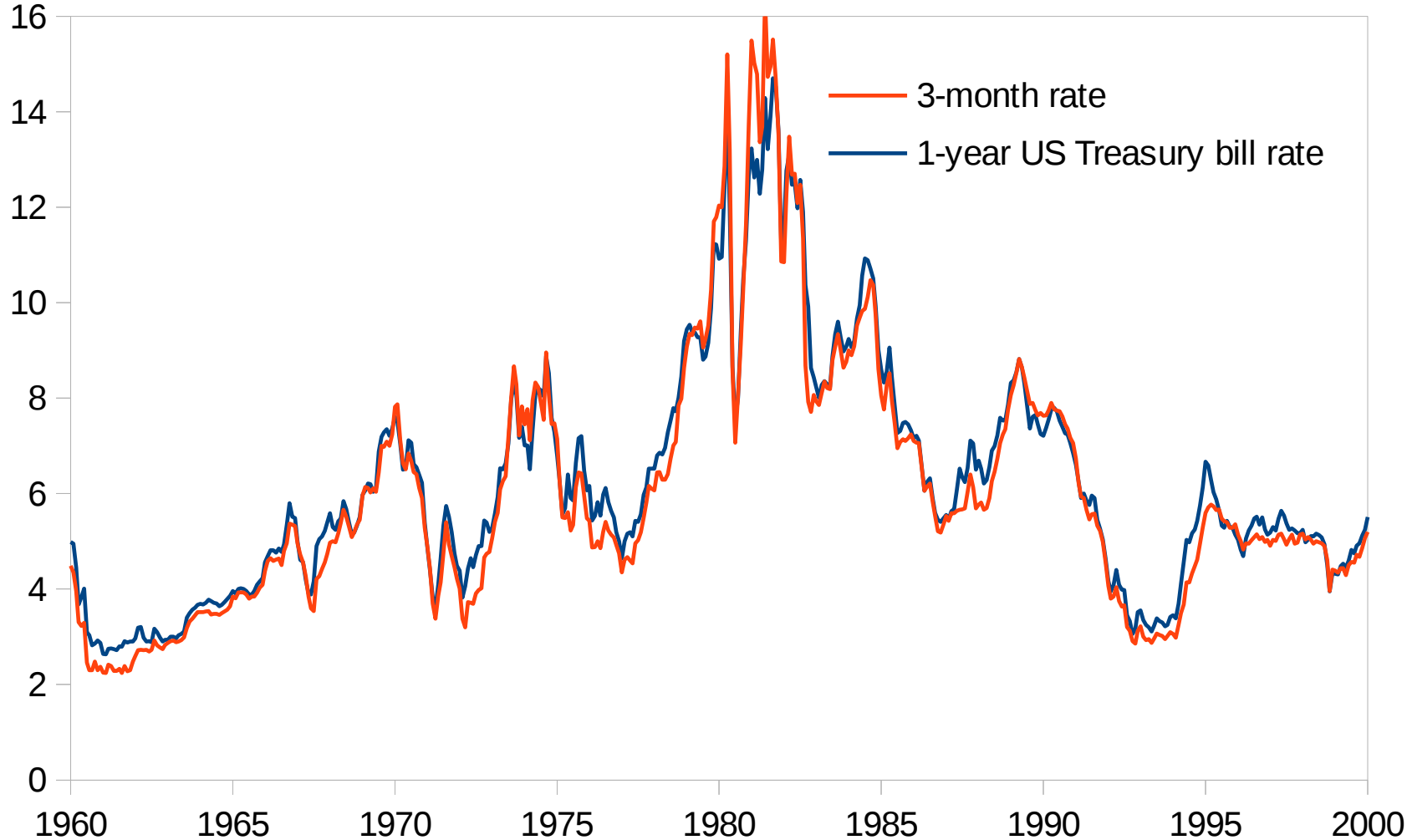
Section 6

Cointegration: a brief review

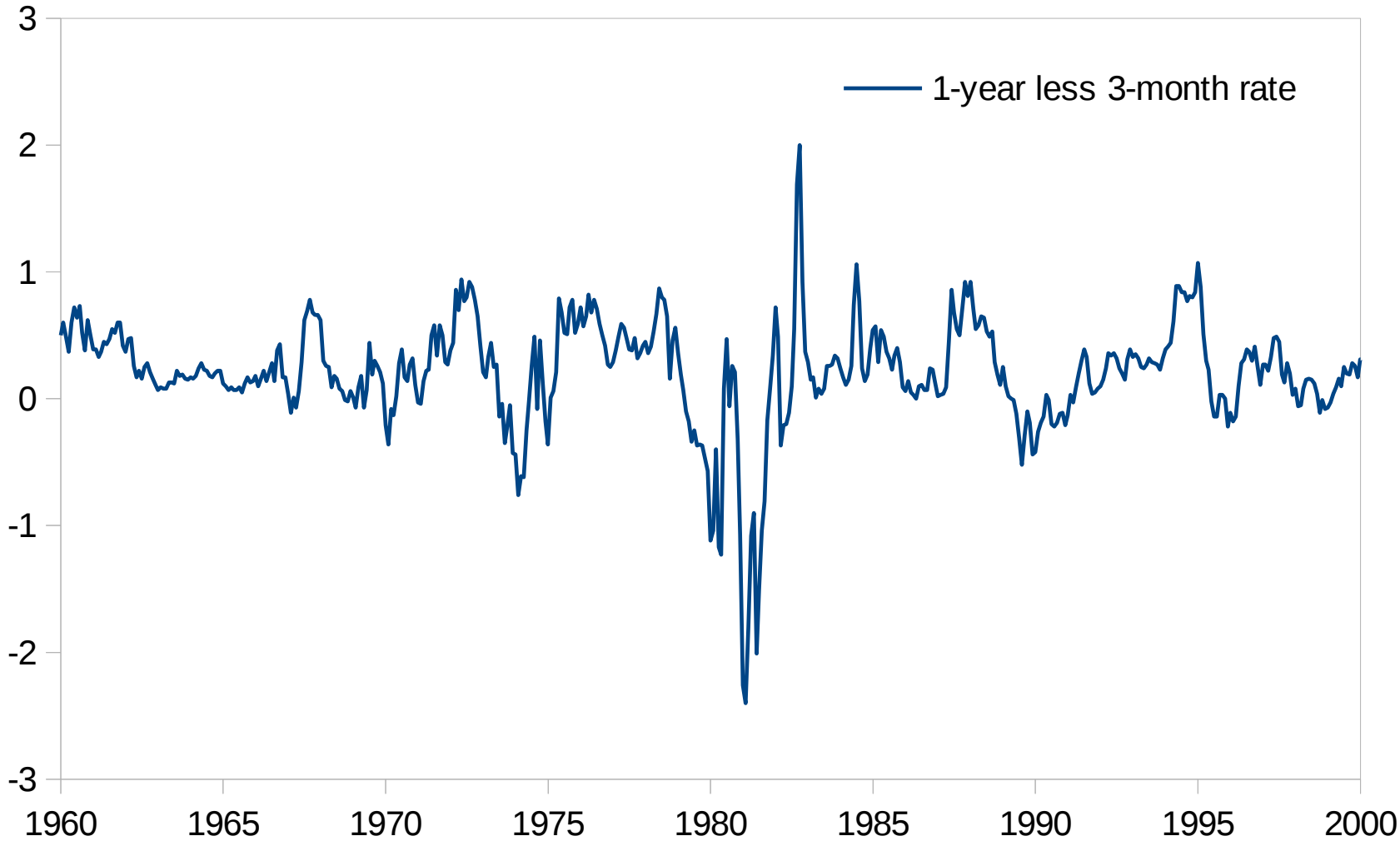
Cointegration

- Characteristics of time series data in much of macroeconomics and finance:
  1. individual series (even after deterministic detrending) are highly persistent (‘randomly wandering‘ / ‘nonstationary’ / ‘integrated’)
  2. but certain linear combinations remain stable (‘stationary’) over time; describe the ‘long-run equilibrium’ relationships between the series
- Mathematically modelled as (common) stochastic trends

Example: US Treasury bill rates



Example: term spread



Orders of integration

- $z_t \sim I(0)$ [integrated of order 0] if it is stationary (up to initialisation) and has *nonzero* long-run variance

$$\text{lrvar}(z_t) := \sum_{k=-\infty}^{\infty} \text{cov}(z_t, z_{t-k})$$

  - requirement that $\text{lrvar}(z_t) \neq 0$ prevents 'overdifferencing'
  - e.g. $z_t = \Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}$ for ε_t i.i.d., is stationary but *not* $I(0)$
- $z_t \sim I(d)$ [integrated of order d] if $\Delta^d z_t \sim I(0)$
  - loosely speaking, if it can be reduced to stationarity by differencing d times
  - d is the *minimum* amount of differencing needed to yield stationarity
- Here focus on the $I(0)$ / $I(1)$ cases, due to their empirical prominence
- So what are the properties of an $I(1)$ series? Can they exhibit the properties noted previously?

Beveridge–Nelson (BN) decomposition

- Suppose $z_t \sim I(1)$ is p-dimensional, and for $\{u_t\}$ i.i.d.,

$$\Delta z_t = \delta + \sum_{i=0}^{\infty} \Theta_i u_{t-i} \qquad \sum_{i=0}^{\infty} i\|\Theta_i\| < \infty$$

- Multivariate Beveridge–Nelson decomposition (Phillips and Solo, 1992):

$$z_t = z_0 - v_0 + \delta t + \Theta(1) \sum_{s=1}^{t} u_s + v_t \qquad \underbrace{\Theta(1)}_{\text{rank } q} := \sum_{i=0}^{\infty} \Theta_i$$

  reduces z_t to an initial condition and three components:

  1. δt: deterministic linear trend
  2. $\Theta(1) \sum_{s=1}^{t} u_s$: q common random walks ('stochastic trends')
  3. $v_t = \sum_{i=0}^{\infty} \tilde{\Theta}_i u_{t-i}$, for $\tilde{\Theta}_i := -\sum_{j=i+1}^{\infty} \Theta_j$, is a stationary linear process
     - also termed the 'common trends representation' (Stock and Watson, 1988)

- Allows us to match the drift, random wandering, high persistence, and transitory fluctuations of many economic time series

Cointegration

$$z_t = z_0 - v_0 + \delta t + \Theta(1) \sum_{s=1}^{t} u_s + v_t$$

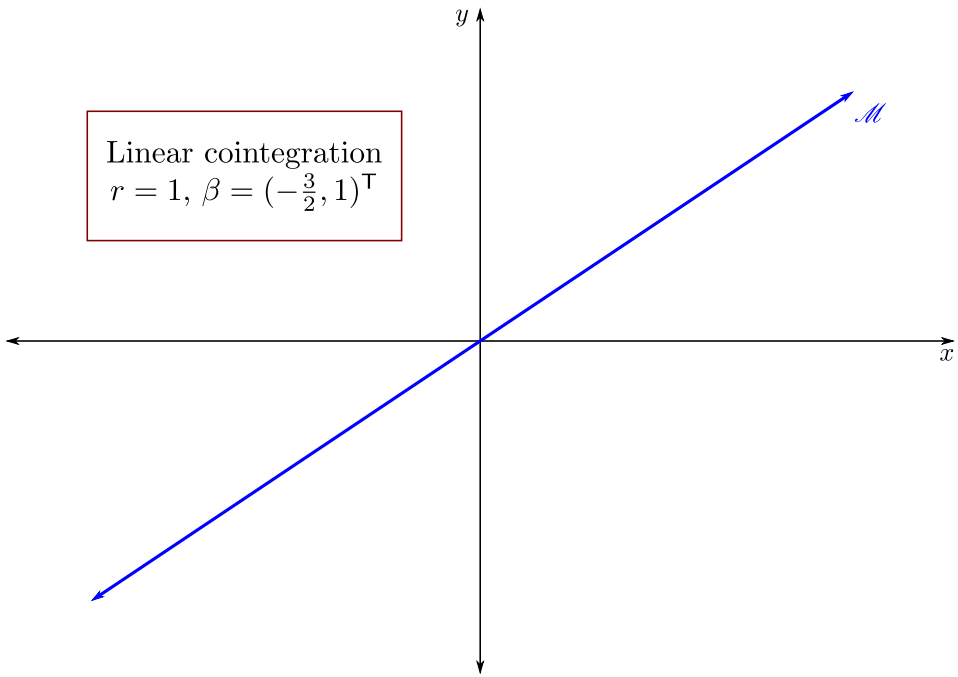
- Suppose $\operatorname{rk} \Theta(1) = q < p$ and $\delta \in \operatorname{sp} \Theta(1)$
  - then there exists $\beta \in \mathbb{R}^{p \times r}$ such that $\beta^\top \Theta(1) = 0$, and so

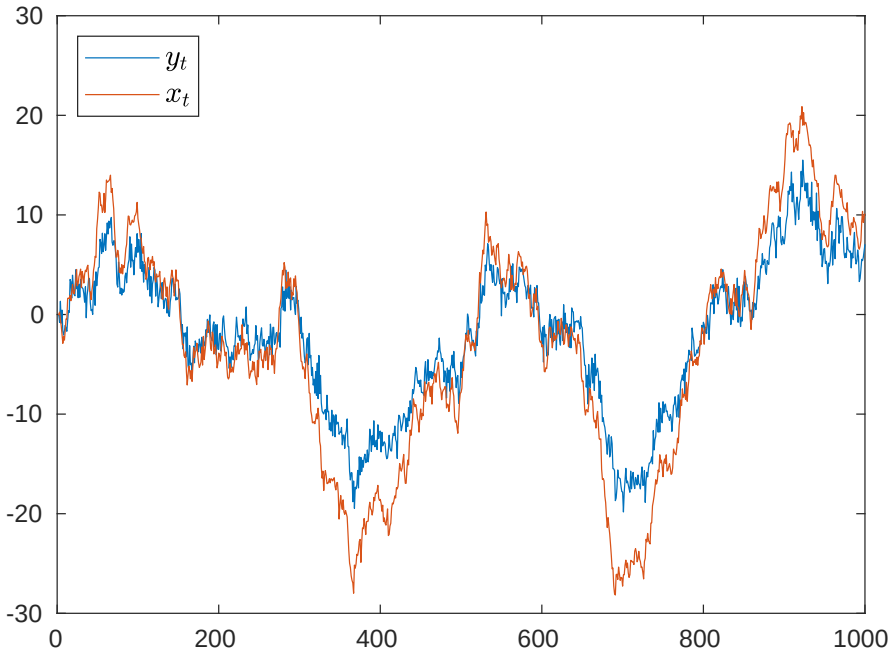
$$\beta^\top z_t = \beta^\top (z_0 - v_0) + \beta^\top v_t$$

  - though z_t wanders around $\mathbb{R}^p$, $\beta^\top z_t$ does not fluctuate too far from its mean
- We say $\{z_t\}$ is *cointegrated* (Engle and Granger, 1987)
  - with *cointegrating rank* $r = \operatorname{rk} \beta = p - q$
  - and *cointegrating space* equal to the span of β
- β is only 'identified' up to its column span, and so may be normalised in a convenient way, e.g. as

$$\beta = \begin{bmatrix} I_r \\ -A^\top \end{bmatrix} \qquad A \in \mathbb{R}^{r \times q}$$

Cointegration





- $z_t = (y_t, x_t)$ is attracted to the one-dimensional linear subspace:

$$\mathscr{M} = \operatorname{sp} \beta_\perp = \operatorname{sp} \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}$$

Section 7

Cointegration in a linear SVAR

Cointegration in a linear SVAR

- p-dimensional SVAR(k): $\Phi(L)z_t = c + \varepsilon_t$, i.e.

$$\Phi_0 z_t = c + \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t$$

- Equivalent VECM form: for $\Pi := -\Phi(1) = \sum_{i=1}^{k} \Phi_i - \Phi_0$

$$\Phi_0 \Delta z_t = c + \Pi z_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta z_{t-i} + \varepsilon_t$$

- Granger–Johansen representation theorem (GJRT):
  - provides conditions under which this model generates cointegrated $I(1)$ series
  - potentially provides useful identifying restrictions
  - see Johansen (1995, Ch. 4) or Hansen (2005) for an exposition
- We only consider the $I(0)/I(1)$ case; higher orders of integration possible

The GJRT in a linear SVAR

$$\Phi_0 z_t = \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t \qquad \Phi(\lambda) = \Phi_0 - \sum_{i=1}^{k} \Phi_i \lambda^i$$

($\mathcal{J}$.i) Φ_0 is invertible

($\mathcal{J}$.ii) Rank: $\operatorname{rk}\Phi(1) = r = p - q$; so $\exists \alpha, \beta \in \mathbb{R}^{p\times r}$ such that

$$\Pi = -\Phi(1) = \alpha\beta^\top \qquad \operatorname{rk}\alpha = \operatorname{rk}\beta = r$$

($\mathcal{J}$.iii) Roots: $\Phi(\lambda)$ has q roots at unity; all others outside unit circle

- Let $\alpha_\perp, \beta_\perp \in \mathbb{R}^{p\times q}$ have rank q, with $\alpha_\perp^\top \alpha = \beta_\perp^\top \beta = 0$. Then:

$$z_t = \underbrace{\beta_\perp(\alpha_\perp^\top \mathrm{H} \beta_\perp)^{-1}\alpha_\perp^\top}_{\text{rank} = q} \sum_{s=1}^{t} \varepsilon_s + \underbrace{M v_t}_{\text{stable}} + \underbrace{a_0}_{\text{init.}} \sim I(1)$$

- $\mathrm{H} = \Phi_0 - \sum_{i=1}^{k-1} \Gamma_i$; $\{v_t\}$ is stationary (up to initialisation)
- Cointegrating rank is $r = p - q$; cointegrating space is $\operatorname{sp}\beta$

The GJRT in a linear SVAR: outline of proof

- Suppose $k = 1$ and $\Phi_0 = I_p$ for simplicity

$$\begin{aligned} z_t = \Phi_1 z_{t-1} + \varepsilon_t \implies \Delta z_t &= (\Phi_1 - I_p) z_{t-1} + \varepsilon_t \\ &= \Pi z_{t-1} + \varepsilon_t \\ &= \alpha\beta^\top z_{t-1} + \varepsilon_t \end{aligned}$$

- Want to show Δz_t is stationary $\implies$ requires $\beta^\top z_{t-1}$ stationary
- Implied autoregressive representation for $\xi_t := \beta^\top z_t$:

$$\begin{aligned} \beta^\top \Delta z_t = \beta^\top \alpha\beta^\top z_{t-1} + \beta^\top \varepsilon_t & \\ \implies \beta^\top z_t = (I_r + \beta^\top \alpha)\beta^\top z_{t-1} + \beta^\top \varepsilon_t & \\ \implies \xi_t = (I_r + \beta^\top \alpha)\xi_{t-1} + \beta^\top \varepsilon_t & \end{aligned}$$

- Is this a stable process, i.e. does $\rho(I_r + \beta^\top \alpha) < 1$?

The GJRT in a linear SVAR: outline of proof

- Recall from $(\mathcal{J}.\text{ii})$:

$$\alpha\beta^\top = \Pi = \Phi_1 - I_p \implies \Phi_1 = I_p + \alpha\beta^\top$$

  so Φ_1 has *at least* q unit eigenvalues
- $(\mathcal{J}.\text{iii})$ says Φ_1 has exactly q unit eigenvalues, r inside the unit circle
- Relate $\alpha\beta^\top \in \mathbb{R}^{p\times p}$ to $\beta^\top\alpha \in \mathbb{R}^{r\times r}$ via the elementary identity:

$$\mu^{p-r}\det(\mu I_r - \beta^\top\alpha) = \det(\mu I_p - \alpha\beta^\top)$$

- Equivalently, taking $\mu = \lambda - 1$:

$$(\lambda-1)^{p-r}\det[\lambda I_r - (I_r + \beta^\top\alpha)] = \det[\lambda I_p - (I_p + \alpha\beta^\top)] = \det(\lambda I_p - \Phi_1)$$

  - l.h.s. is a polynomial with $p - r = q$ roots at unity
  - remaining r roots equal the stable eigenvalues of Φ_1
  - these are exactly the r eigenvalues of $I_r + \beta^\top\alpha$!

The GJRT in a linear SVAR: outline of proof

$$\Delta z_s = \alpha\beta^\top z_{s-1} + \varepsilon_s = \alpha\xi_{s-1} + \varepsilon_s$$

- We also need to extract q common trends
- Premultiply by $\alpha_\perp \in \mathbb{R}^{p\times q}$, where $\alpha_\perp^\top\alpha = 0$:

$$\alpha_\perp^\top \Delta z_s = \alpha_\perp^\top(\alpha\xi_{s-1} + \varepsilon_s) = \alpha_\perp^\top \varepsilon_s$$

- Cumulate from $s = 1$ up to $s = t$:

$$\alpha_\perp^\top z_t = \alpha_\perp^\top z_0 + \sum_{s=1}^{t} \alpha_\perp^\top \Delta z_s = \alpha_\perp^\top z_0 + \alpha_\perp^\top \sum_{s=1}^{t} \varepsilon_s$$

The GJRT in a linear SVAR: outline of proof

- We have obtained the representation:

$$\begin{bmatrix} \alpha_\perp^\top \\ \beta^\top \end{bmatrix} z_t = \begin{bmatrix} \alpha_\perp^\top z_0 + \alpha_\perp^\top \sum_{s=1}^t \varepsilon_s \\ \xi_t \end{bmatrix}$$

- Is the r.h.s. invertible?
  - because $\rho(I_r + \beta^\top \alpha) < 1$, we must have $\mathrm{rk}\, \beta^\top \alpha = r$
  - in this case, elementary matrix algebra yields that

$$\mathrm{rk}[\alpha_\perp, \beta] = \mathrm{rk}[\beta_\perp, \alpha] = p$$

  - moreover,

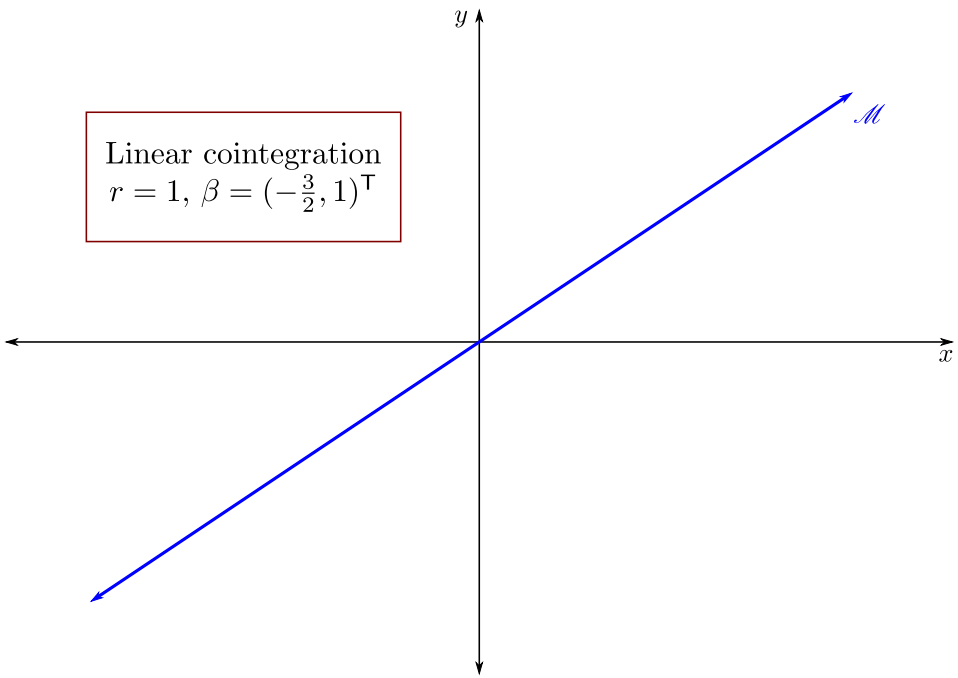
$$I_p = \beta_\perp(\alpha_\perp^\top \beta_\perp)^{-1}\alpha_\perp^\top + \alpha(\beta^\top \alpha)^{-1}\beta^\top =: P_{\beta_\perp} + P_\alpha$$

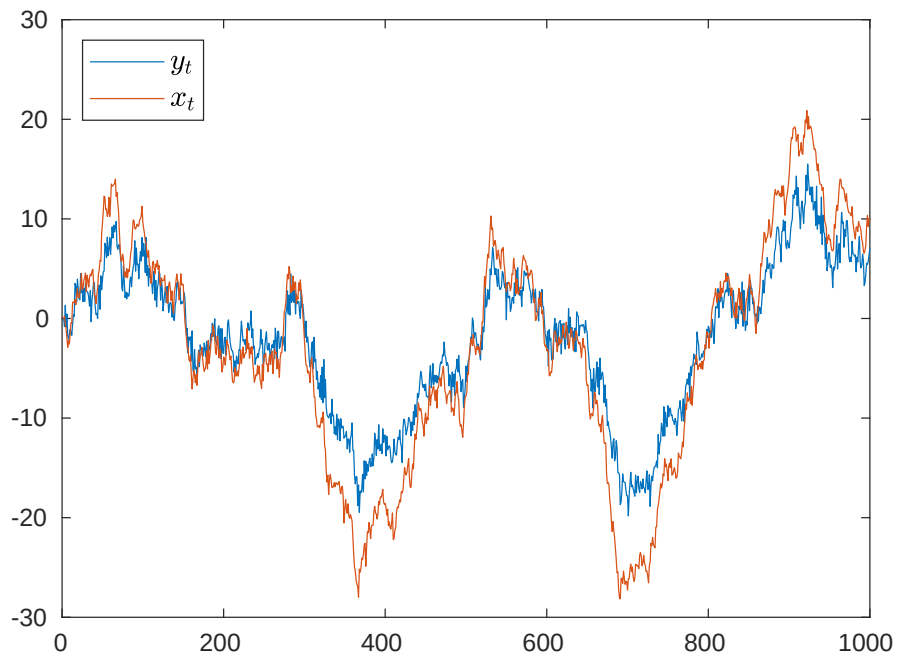
- As a neat alternative to directly inverting the above, we may decompose

$$z_t = (P_{\beta_\perp} + P_\alpha) z_t = P_{\beta_\perp} z_0 + P_{\beta_\perp} \sum_{s=1}^t \varepsilon_s + \alpha(\beta^\top \alpha)^{-1} \xi_t$$

The GJRT in a linear SVAR

$$z_t = \beta_\perp(\alpha_\perp^\top \mathsf{H}\beta_\perp)^{-1}\alpha_\perp^\top \sum_{s=1}^{t}\varepsilon_s + \underbrace{Mv_t}_{\text{stable}} + \underbrace{a_0}_{\text{init.}}$$





The GJRT in a linear SVAR: key implications

- Recall that parameters of the SVAR are identified only up to some $Q \in \mathbb{O}(p)$

$$\Phi_0 z_t = c + \sum_{i=1}^{k} \Phi_i z_{t-i} + \varepsilon_t$$

- Preceding is observationally equivalent to

$$\tilde{\Phi}_0 z_t = \tilde{c} + \sum_{i=1}^{k} \tilde{\Phi}_i z_{t-i} + Q\varepsilon_t$$

  - $\tilde{\Phi}_0$ is lower triangular, with positive diagonal entries
  - $(\tilde{c}, \{\tilde{\Phi}_i\}_{i=0}^{k})$ is identified, $Q \in \mathbb{O}(p)$ is not

- We can use the GJRT to obtain restrictions on Q ...

The GJRT in a linear SVAR: key implications

$$\tilde{\Phi}_0 z_t = \tilde{c} + \sum_{i=1}^{k} \tilde{\Phi}_i z_{t-i} + Q\varepsilon_t$$

- Apply the GJRT to this system, with now

$$\Pi := -\tilde{\Phi}(1) = \sum_{i=1}^{k} \tilde{\Phi}_i - \tilde{\Phi}_0 = \alpha\beta^\top$$

  - yields the representation

$$z_t = \underbrace{\beta_\perp(\alpha_\perp^\top \tilde{\mathsf{H}}\beta_\perp)^{-1}\alpha_\perp^\top}_{\text{identified}} Q \sum_{s=1}^{t} \varepsilon_s + Mv_t + a_0$$

  - structural shocks may have a *permanent* effect only via $\alpha_\perp^\top Q \sum_{s=1}^{t} \varepsilon_s$

- Since $\operatorname{rk}\alpha_\perp = q$, *up to* $r = p - q$ structural shocks may have purely transitory effects

The GJRT in a linear SVAR: key implications

- Wlog, order $\varepsilon_t = (\varepsilon_{(1),t}^\top, \varepsilon_{(2),t}^\top)^\top$ so that (say) r shocks in $\varepsilon_{(1),t}$ have only transitory effects:

$$0 = \alpha_\perp^\top Q \sum_{s=1}^{t} \begin{bmatrix} \varepsilon_{(1),s} \\ 0 \end{bmatrix} \implies 0 = \alpha_\perp^\top Q \begin{bmatrix} I_r \\ 0_{q\times r} \end{bmatrix}$$

- We obtain qr long-run identifying restrictions (Blanchard and Quah, 1989; King, Plosser, Stock, and Watson, 1991; Kilian and Lütkepohl, 2017, Sec. 10.2)

The GJRT in a linear SVAR: key implications

$$z_t = \underbrace{\beta_\perp(\alpha_\perp^\top \mathsf{H}\beta_\perp)^{-1}\alpha_\perp^\top}_{=:P_{\beta_\perp}} \sum_{s=1}^{t} \varepsilon_s + \underbrace{Mv_t}_{\text{stable}} + \underbrace{a_0}_{\text{init.}}$$

- Implied limit theory for $n^{-1/2} z_{\lfloor n\lambda \rfloor}$: vital for asymptotic inference
- We have, in particular,

$$\begin{aligned} Z_n(\lambda) := n^{-1/2} z_{\lfloor n\lambda \rfloor} &= P_{\beta_\perp} n^{-1/2} \sum_{s=1}^{\lfloor n\lambda \rfloor} \varepsilon_s + n^{-1/2}(Mv_{\lfloor n\lambda \rfloor} + a_0) \\ &= P_{\beta_\perp} E_n(\lambda) + o_p(1) \\ &\rightsquigarrow P_{\beta_\perp} E(\lambda) =: Z(\lambda) \end{aligned}$$

- $Z(\lambda)$ is a p-dimensional Brownian motion obliquely projected onto the q-dimensional subspace,

$$\mathscr{M} = \operatorname{sp} \beta_\perp = \ker \Pi$$

the attractor set for $\{z_t\}$

Section 8

The GJRT in a nonlinear SVAR

Nonlinear VECM representation

$$f_0(z_t) = \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

- To extend the GJRT to this setting: need a VECM representation to isolate the 'stochastic trend' and 'weakly dependent' components
- Here additive time-separability helps greatly: e.g. when $k = 2$

$$\begin{aligned}
\Delta f_0(z_t) &= -f_0(z_{t-1}) + f_1(z_{t-1}) + f_2(z_{t-2}) + \varepsilon_t \\
&= (f_1 + f_2 - f_0)(z_{t-1}) - f_2(z_{t-1}) + f_2(z_{t-2}) + \varepsilon_t \\
&= (f_1 + f_2 - f_0)(z_{t-1}) - \Delta f_2(z_{t-1}) + \varepsilon_t \\
&=: \pi(z_{t-1}) + \Delta g_1(z_{t-1}) + \varepsilon_t
\end{aligned}$$

- More generally,

$$\Delta f_0(z_t) = \pi(z_{t-1}) + \sum_{i=1}^{k-1} \Delta g_i(z_{t-i}) + \varepsilon_t$$

$$\pi(z) := \sum_{i=1}^{k} f_i(z) - f_0(z) \qquad g_j(z) := -\sum_{i=j+1}^{k} f_i(z)$$

The GJRT in a nonlinear SVAR(1)

$$f_0(z_t) = f_1(z_{t-1}) + \varepsilon_t \qquad \Longleftrightarrow \qquad \Delta f_0(z_t) = \pi(z_{t-1}) + \varepsilon_t$$

($\mathcal{M}$.i) $f_0 : \mathbb{R}^p \to \mathbb{R}^p$ is a homeomorphism

($\mathcal{M}$.ii) there exists an $\alpha \in \mathbb{R}^{p\times r}$ with $\operatorname{rk}\alpha = r$, and $\theta : \mathbb{R}^p \to \mathbb{R}^r$ such that

$$\pi(z) = \alpha\theta(z)$$

($\mathcal{M}$.iii) for every $z, z' \in \mathbb{R}^p$, $\exists \beta \in \mathcal{B}$ compact such that

$$\theta(z) - \theta(z') = \beta^\top (z - z') \qquad [\text{if } f_0 = I_p]$$

and $\rho_{\mathrm{JSR}}\{I_r + \beta^\top \alpha \mid \beta \in \mathcal{B}\} \leq \overline{\rho} < 1$

- Then $\xi_t := \theta(z_t)$ is 'stable', and there exists a homeomorphism $\chi : \mathbb{R}^p \to \mathbb{R}^p$ such that

$$z_t = \chi^{-1}\begin{bmatrix} \alpha_\perp^\top f_0(z_0) + \alpha_\perp^\top \sum_{s=1}^t \varepsilon_s \\ \xi_t \end{bmatrix}$$

$(\mathcal{M}.\text{i})$: f_0 is a homeomorphism

$$f_0(z_t) = f_1(z_{t-1}) + \varepsilon_t$$

- Can rewrite the model in terms of $z_t^* := f_0(z_t)$ and $f_i^* := f_i \circ f_0^{-1}$:

$$\begin{aligned} z_t^* = f_0(z_t) &= f_1(z_{t-1}) + \varepsilon_t \\ &= f_1\{f_0^{-1}[f_0(z_{t-1})]\} + \varepsilon_t \\ &= (f_1 \circ f_0^{-1})(z_{t-1}^*) + \varepsilon_t \\ &= f_1^*(z_{t-1}^*) + \varepsilon_t \end{aligned}$$

- So we can focus on the simpler case where $f_0 = I_p$

$(\mathcal{M}.\text{ii})$: counterpart of the rank condition

$$\Delta f_0(z_t) = \pi(z_{t-1}) + \varepsilon_t = \alpha\theta(z_{t-1}) + \varepsilon_t$$

$(\mathcal{M}.\text{ii})$ there exists $\alpha \in \mathbb{R}^{p\times r}$ with $\operatorname{rk}\alpha = r$, and $\theta : \mathbb{R}^p \to \mathbb{R}^r$ such that

$$\pi(z) = \alpha\theta(z)$$

- We can isolate the common trends similarly to the linear case:

$$\alpha_\perp^\top \Delta f_0(z_t) = \alpha_\perp^\top \varepsilon_t \implies \alpha_\perp^\top f_0(z_t) = \alpha_\perp^\top f_0(z_0) + \alpha_\perp^\top \sum_{s=1}^{t} \varepsilon_s$$

- Define $\psi : \mathbb{R}^p \to \mathbb{R}^q$ as the map that extracts the common trends

$$\psi(z) \coloneqq \alpha_\perp^\top f_0(z)$$

- In a CKSVAR, this condition can be relaxed; it corresponds to case (ii) in Duffy et al. (2025)

($\mathcal{M}$.iii): counterpart of the root condition

- Under ($\mathcal{M}$.i), we may effectively suppose $f_0 = I_p$ without loss of generality

$$\Delta z_t = \pi(z_{t-1}) + \varepsilon_t = \alpha\theta(z_{t-1}) + \varepsilon_t =: \alpha\xi_{t-1} + \varepsilon_t$$

- Want the 'equilibrium errors' $\xi_t := \theta(z_t)$ to be of smaller order than the common trends
- Suppose there exists a $\beta_t \in \mathbb{R}^{p\times r}$, depending on z_t and z_{t-1}, such that

$$\Delta\xi_t = \theta(z_t) - \theta(z_{t-1}) = \beta_t^\top(z_t - z_{t-1}) = \beta_t^\top \Delta z_t$$

- Leads to a *time-varying* autoregression for ξ_t:

$$\Delta\xi_t = \beta_t^\top[\alpha\xi_{t-1} + \varepsilon_t] \implies \xi_t = (I_r + \beta_t^\top\alpha)\xi_{t-1} + \beta_t^\top\varepsilon_t$$

- [In the linear model, $\theta(z) = \beta^\top z$, $\xi_t = \beta^\top z_t$, and $\beta_t = \beta$ for all t.]

$(\mathcal{M}.\text{iii})$: counterpart of the root condition

$$\xi_t = (I_r + \beta_t^\top \alpha)\xi_{t-1} + \beta_t^\top \varepsilon_t \qquad\qquad \Delta\theta(z_t) = \beta_t^\top \Delta z_t$$

- Let $\mathcal{B} \subset \mathbb{R}^{p\times r}$ be such that:

$$\forall z, z' \in \mathbb{R}^p, \exists \beta \in \mathcal{B} \text{ s.t. } \theta(z) - \theta(z') = \beta^\top (z - z')$$

- [$\mathcal{B}$ is bounded $\iff$ $\theta(z)$ is Lipschitz]
- Want $\{\xi_t\}$ to follow a stable autoregression: need to control:

$$\prod_{s=1}^{t} (I_r + \beta_s^\top \alpha), \qquad \{\beta_s\}_{s=1}^t \subset \mathcal{B}.$$

- In a linear model: $\mathcal{B} = \{\beta\}$;
  - control follows from $\rho(I_r + \beta^\top \alpha) < 1$
  - implied by (and equivalent to!) the root condition
- So here instead we impose a condition on the *joint* spectral radius

Interlude: the joint spectral radius

- The *joint spectral radius* of a bounded set $\mathcal{A} \subset \mathbb{R}^{m\times m}$ of matrices:

$$\rho_{\mathrm{JSR}}(\mathcal{A}) := \limsup_{t\to\infty} \sup_{B\in\mathcal{A}^t} \rho(B)^{1/t}$$

  for $\mathcal{A}^t := \{\prod_{s=1}^t M_s \mid M_s \in \mathcal{A}\}$ the t-fold products of matrices in $\mathcal{A}$.

- In a dynamical system, if $\rho_{\mathrm{JSR}}(\mathcal{A}) < 1$,

$$w_t = A_t w_{t-1} + v_t \qquad\qquad A_t \in \mathcal{A}$$

  (a) if $v_t := 0\ \forall t$, then $w_t \to 0$, for every w_0
  (b) if $\|w_0\|_{2+\delta}$ and $\sup_{t\in\mathbb{N}}\|v_t\|_{2+\delta}$ finite, then $\max_{1\le t\le n}\|w_t\| = o_p(n^{1/2})$

- Used previously in the analysis of nonlinear autoregressive models, to establish ergodicity (e.g. Liebscher, 2005; Saikkonen, 2008, etc.)
- Substantial literature on computing approximate upper bounds for the JSR (see Duffy et al., 2023)

$(\mathcal{M}.\text{iii})$: counterpart of the root condition

$$\xi_t = (I_r + \beta_t^\top \alpha)\xi_{t-1} + \beta_t^\top \varepsilon_t \qquad\qquad \beta_t \in \mathcal{B}$$

$(\mathcal{M}.\text{iii})$ for every $z, z' \in \mathbb{R}^p$, $\exists \beta \in \mathcal{B}$ compact such that

$$\theta(z) - \theta(z') = \beta^\top (z - z') \qquad [\text{if } f_0 = I_p]$$

and $\rho_{\text{JSR}}\{I_r + \beta^\top \alpha \mid \beta \in \mathcal{B}\} \leq \overline{\rho} < 1$

- When $f_0 \neq I_p$, amend to

$$(\theta \circ f_0^{-1})(z) - (\theta \circ f_0^{-1})(z') = \beta^\top (z - z')$$

- This is the only other condition we need!

The GJRT in a nonlinear SVAR(1)

$$\Delta f_0(z_t) = \pi(z_{t-1}) + \varepsilon_t = \alpha\theta(z_{t-1}) + \varepsilon_t$$

- We have now shown that, for $\chi : \mathbb{R}^p \to \mathbb{R}^p$ defined as

$$\chi(z_t) := \begin{bmatrix} \psi(z_t) \\ \theta(z_t) \end{bmatrix} = \begin{bmatrix} \alpha_\perp^\top f_0(z_t) \\ \theta(z_t) \end{bmatrix} = \begin{bmatrix} \alpha_\perp^\top f_0(z_0) + \alpha_\perp^\top \sum_{s=1}^t \varepsilon_s \\ \xi_t \end{bmatrix} \quad \begin{matrix} q \text{ trends} \\ r \text{ wk. dep.} \end{matrix}$$

  - q common stochastic trends: $\alpha_\perp^\top \sum_{s=1}^t \varepsilon_s$
  - $\max_{t \le n} \|\xi_t\| = o_p(n^{1/2})$ is of smaller order (though *not* stationary)
- Is χ invertible? Under $(\mathcal{M}.\text{i})$–$(\mathcal{M}.\text{iii})$ it is a homeomorphism; hence

$$z_t = \chi^{-1} \begin{bmatrix} \alpha_\perp^\top f_0(z_0) + \alpha_\perp^\top \sum_{s=1}^t \varepsilon_s \\ \xi_t \end{bmatrix}$$

The GJRT: the general result

$$\Delta f_0(z_t) = \pi(z_{t-1}) + \sum_{i=1}^{k-1} \Delta g_i(z_{t-i}) + \varepsilon_t \qquad g_j(z) := -\sum_{i=j+1}^{k} f_i(z)$$

- Handle $k \geq 2$ via a companion form (cf. Beare, Seo, and Seo, 2017)
- This would not be intrinsically interesting, except that the nonlinearity inhibits writing this in terms of the usual state vector

$$\vec{z}_{t-1} = (z_{t-1}^\top, z_{t-2}^\top, \ldots, z_{t-k}^\top)^\top$$

- Additive separability allows us to transfer all the nonlinearity to the first p components of the state vector, defined instead as

$$\boldsymbol{z}_{t-1} := (z_{t-1}^\top, \zeta_{1,t-2}^\top, \ldots, \zeta_{k-1,t-2}^\top)^\top \qquad \zeta_{j,t} := \sum_{i=j}^{k-1} g_i(z_{t-i+j})$$

The GJRT: the general result

$$\Delta f_0(z_t) = c + \pi(z_{t-1}) + \sum_{i=1}^{k-1} \Delta g_i(z_{t-i}) + \varepsilon_t$$

- $(\mathcal{M}.\text{i})$ and $(\mathcal{M}.\text{ii})$ are exactly as above; in particular we only require $\pi(z) = \alpha\theta(z)$ and $c = \alpha\mu$
- $(\mathcal{M}.\text{iii})$ takes an augmented form, to account for higher-order dynamics

- $\{\boldsymbol{\xi}_t\}$ follows a stable autoregression, and

$$z_t = \chi^{-1}\left(\begin{bmatrix} \alpha_\perp^\top \overline{h}(\vec{z}_0) + \alpha_\perp^\top \sum_{s=1}^{t} \varepsilon_s \\ -\mu \end{bmatrix} + S_\chi^\top \boldsymbol{\xi}_t\right)$$

- χ is a homeomorphism, and

$$\chi(z) = \begin{bmatrix} \psi(z) \\ \theta(z) \end{bmatrix} \qquad \psi(z) = \alpha_\perp^\top \left[f_0(z) - \sum_{i=1}^{k-1} g_i(z) \right]$$

Long-run identifying restrictions

$$\Delta f_0(z_t) = c + \pi(z_{t-1}) + \sum_{i=1}^{k-1} \Delta g_i(z_{t-i}) + \varepsilon_t$$

- Recall parameters are identified up to an unknown $Q \in \mathbb{O}(p)$
- Apply the nonlinear GJRT to the (observationally equivalent) system

$$\Delta \tilde{f}_0(z_t) = \tilde{c} + \tilde{\pi}(z_{t-1}) + \sum_{i=1}^{k-1} \Delta \tilde{g}_i(z_{t-i}) + Q\varepsilon_t$$

  - $D\tilde{f}_0(z^*)$ lower triangular, with positive diagonal, for some (fixed) $z^* \in \mathbb{R}^p$
  - $(\tilde{c}, \tilde{f}_0, \tilde{\pi}, \{\tilde{g}_i\}_{i=1}^{k-1})$ are identified; $Q \in \mathbb{O}(p)$ is not
  - our theorem yields, for $\alpha\theta(z) = \tilde{\pi}(z)$,

$$z_t = \chi^{-1}\left(\begin{bmatrix} \alpha_\perp^\top \overline{h}(\vec{z}_{\tau-1}) + \alpha_\perp^\top Q \sum_{s=\tau}^{t} \varepsilon_s \\ -\mu \end{bmatrix} + S_\chi^\top \xi_t\right)$$

Long-run identifying restrictions

- Framework: [$c = 0$ for simplicity]
  - fix the state $\vec{z}_{t-1} = (z_{t-1}^\top, \ldots, z_{t-k}^\top) = \mathfrak{z} \in \mathbb{R}^{kp}$ at time $t = \tau$
  - one final shock $\varepsilon_\tau = \varepsilon$ occurs; then $\varepsilon_t = 0$ for $t \geq \tau + 1$
- What is the limit of z_t as $t \to \infty$?

$$z_t = \chi^{-1}\left(\begin{bmatrix}\alpha_\perp^\top \overline{h}(\vec{z}_{\tau-1}) + \alpha_\perp^\top Q \sum_{s=\tau}^{t} \varepsilon_s \\ 0\end{bmatrix} + S_\chi^\top \xi_t\right)$$

$$\to \chi^{-1}\left(\begin{bmatrix}\alpha_\perp^\top \overline{h}(\mathfrak{z}) + \alpha_\perp^\top Q\varepsilon \\ 0\end{bmatrix}\right) =: z_\infty(\varepsilon, \mathfrak{z})$$

- $z_\infty(\varepsilon, \mathfrak{z})$ is a steady-state ($z_t = z_{t-1} = \cdots = z_{t-k}$) solution to

$$\Delta \tilde{f}_0(z_t) = \tilde{\pi}(z_{t-1}) + \sum_{i=1}^{k-1} \Delta \tilde{g}_i(z_{t-i}) + Q\varepsilon_t$$

i.e. an element of $\mathscr{M}_0 := \ker \tilde{\pi} = \ker \theta$

Long-run identifying restrictions

$$z_t(\varepsilon;\mathfrak{z}) \to \chi^{-1}\left(\begin{bmatrix}\alpha_\perp^\top \bar{h}(\mathfrak{z}) + \alpha_\perp^\top Q\varepsilon \\ 0\end{bmatrix}\right) =: z_\infty(\varepsilon,\mathfrak{z}) \in \mathscr{M}_0$$

- Shocks ε_t having no permanent effects satisfy

$$z_\infty(\varepsilon,\mathfrak{z}) = z_\infty(0,\mathfrak{z}) \iff \alpha_\perp^\top Q\varepsilon = 0 \iff Q\varepsilon \in \operatorname{sp}\alpha$$

Long-run identifying restrictions

- Suppose we partition

$$\varepsilon_t = \begin{bmatrix}\varepsilon_{(1),t} \\ \varepsilon_{(2),t}\end{bmatrix} \quad \begin{matrix} r \text{ shocks} \\ q \text{ shocks}\end{matrix}$$

- If the r structural shocks in $\varepsilon_{(1),t}$ have only transitory effects, then

$$\alpha_\perp^\top Q \begin{bmatrix} I_r \\ 0_{q\times r}\end{bmatrix} = 0$$

Section 9

Examples and applications

Example: piecewise affine VAR

$$f_0(z_t) = \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

- Specialise to the case where f_i is continuous and of the form

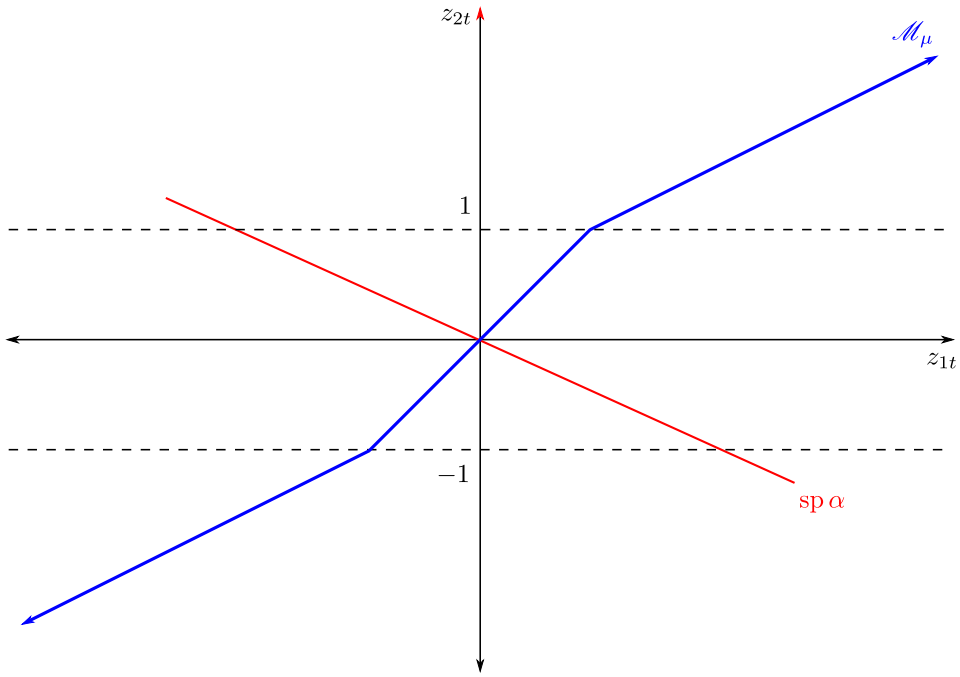
$$f_i(z) = \sum_{\ell=1}^{L} 1\{z \in \mathscr{Z}^{(\ell)}\}(\bar{\phi}_i^{(\ell)} + \Phi_i^{(\ell)} z)$$

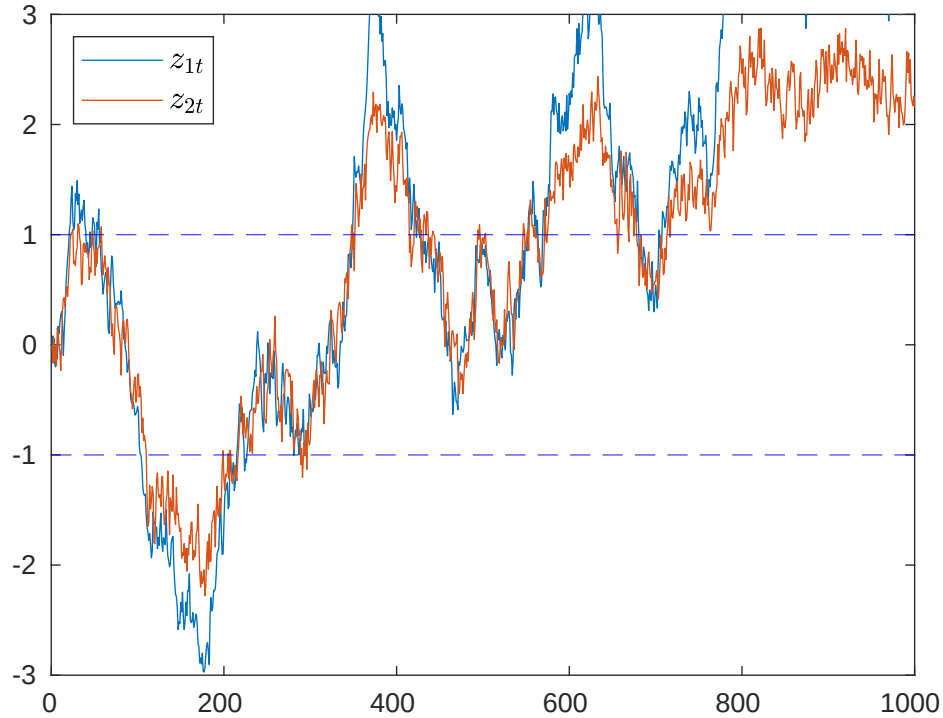
  where $\{\mathscr{Z}^{(\ell)}\}_{\ell=1}^{L}$ are convex and partition $\mathbb{R}^p$
- Elementary example: $p = 2$, $L = 3$, $\nu_1 = -1$, $\nu_2 = 1$

$$\Delta z_t = \alpha \sum_{\ell=1}^{L} 1\{z_{2,t-1} \in (\nu_{\ell-1}, \nu_\ell]\}(\bar{\mu}^{(\ell)} + \beta^{(\ell)\top} z_{t-1}) + \varepsilon_t$$

  with α and $\beta^{(\ell)}$ as follows ...

Example: piecewise affine VAR

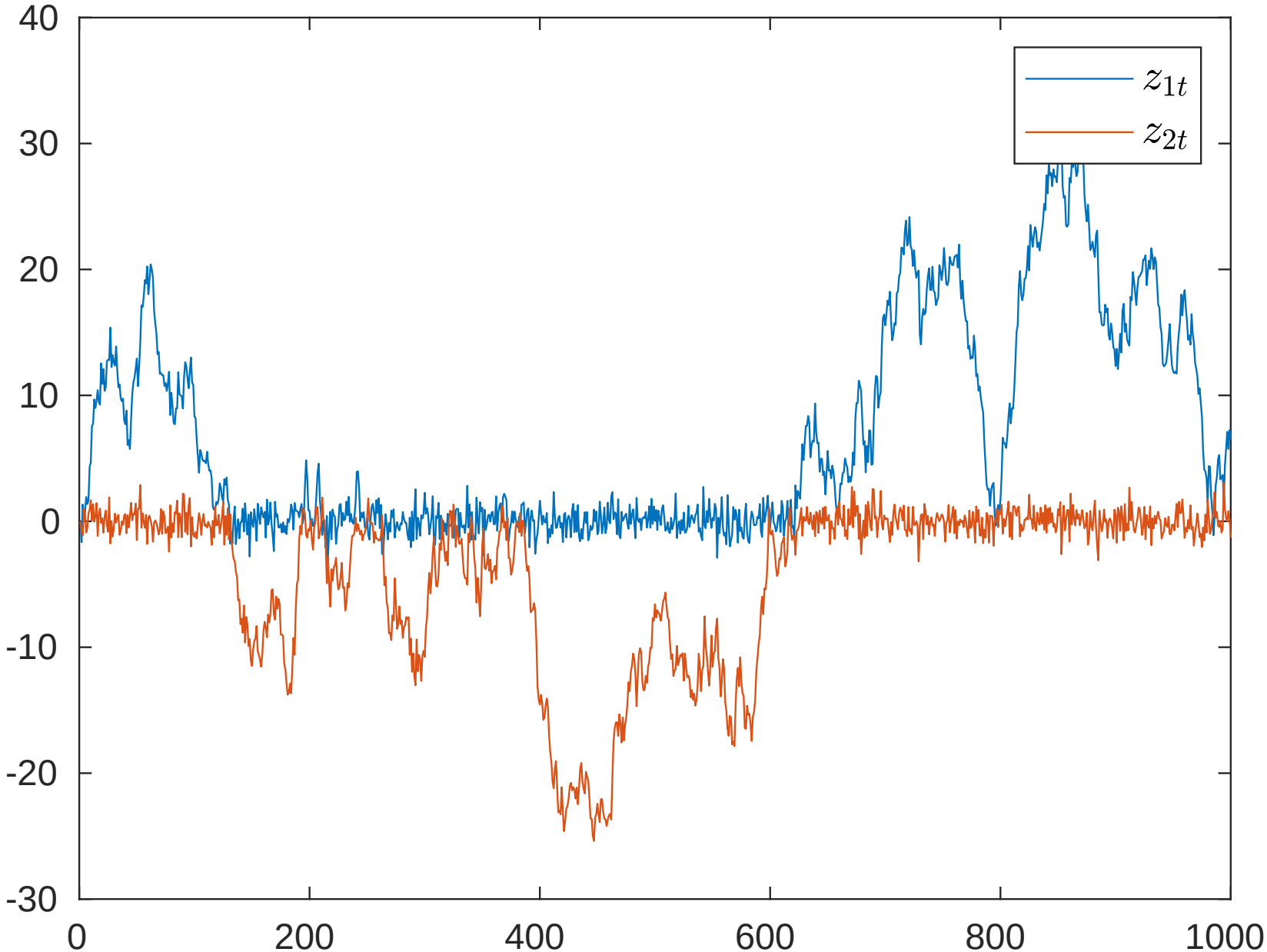




$$\Delta z_t = \underbrace{\alpha \sum_{\ell=1}^{L} 1\{z_{2,t-1} \in (\nu_{\ell-1}, \nu_\ell]\}(\bar{\mu}^{(\ell)} + \beta^{(\ell)\top} z_{t-1})}_{=:\pi(z_{t-1})} + \varepsilon_t$$

$$\alpha = \begin{bmatrix} 1/8 \\ -1/2 \end{bmatrix} \qquad \beta^{(1)} = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix} \qquad \beta^{(2)} = \begin{bmatrix} -1/2 \\ 1/2 \end{bmatrix} \qquad \beta^{(3)} = \begin{bmatrix} -1/2 \\ 1 \end{bmatrix}$$

Example: 'sharing' a common trend



What generated this?

- We should ask: what is the attractor set?
  - z_{1t} wanders freely over $\mathbb{R}^+ \iff z_{2t} \simeq 0$
  - z_{2t} wanders freely over $\mathbb{R}^- \iff z_{1t} \simeq 0$
- Suggests the attractor set

$$\mathscr{M}_0 = \{z \in \mathbb{R}^2 \mid z_1 \geq 0,\ z_2 = 0\} \cup \{z \in \mathbb{R}^2 \mid z_1 = 0,\ z_2 \leq 0\}$$

- A function $\theta : \mathbb{R}^2 \to \mathbb{R}^1$ with $\ker \theta = \mathscr{M}_0$ is

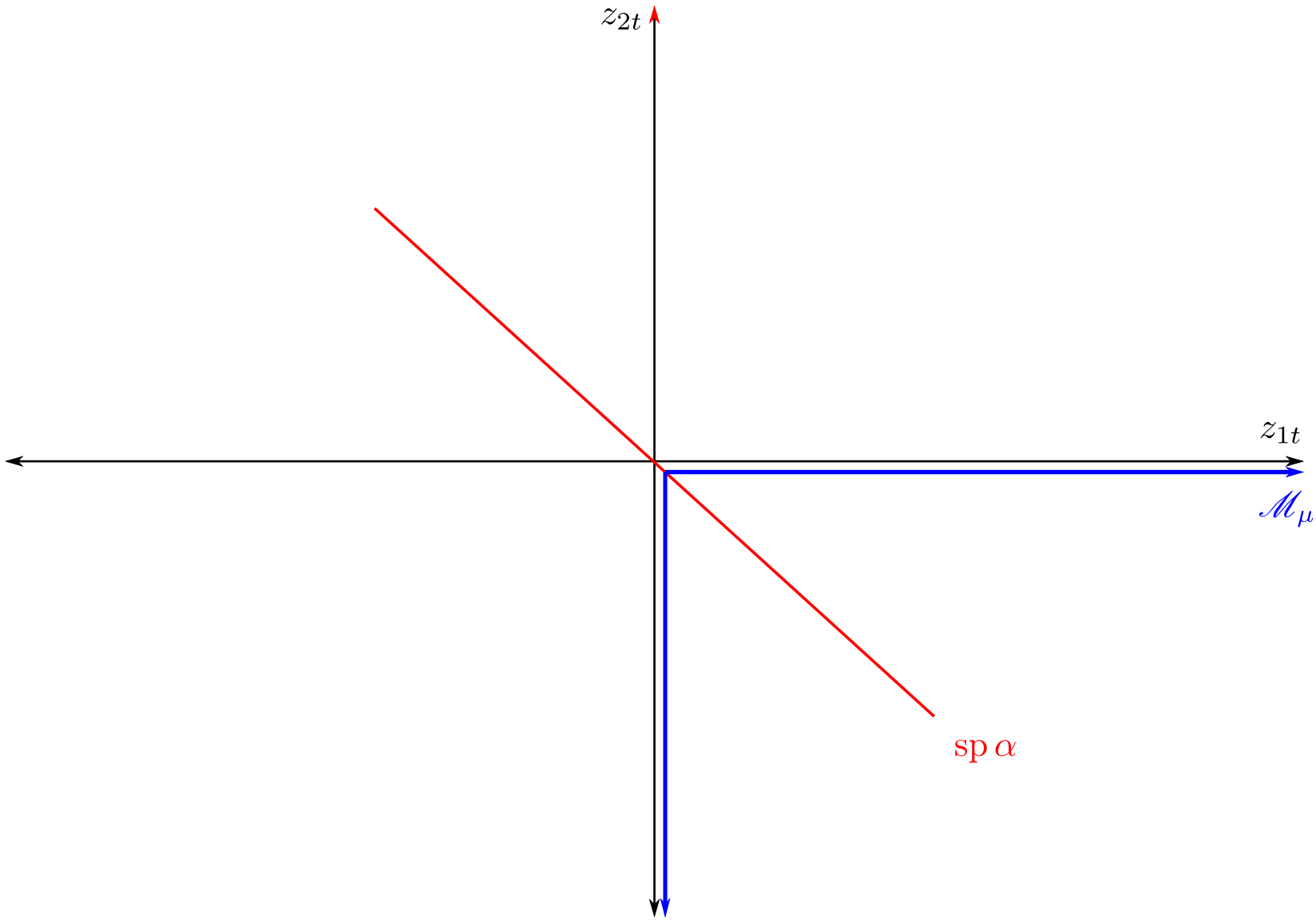
$$\theta(z) = z_2 1\{z_1 + z_2 \geq 0\} - z_1 1\{z_1 + z_2 < 0\}$$

- These trajectories have been generated by simulating

$$\Delta z_t = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \theta(z_{t-1}) + \varepsilon_t$$

which satisfies ($\mathcal{M}$.i)–($\mathcal{M}$.iii)

What generated this?



Applications

1. The long-run Phillips Curve (Ascari, Bonomolo, and Haque, 2023)
   - allow for a long-run relationship between potential output and trend inflation
   - entails a nonlinearly cointegrated VAR model for the trend components
2. Nonlinearity in the global carbon cycle (Haimerl et al, in progress)
   - allow CO_2 uptake of natural carbon sinks to depend on CO_2 levels
   - is there a 'tipping point' at which sinks become markedly less effective?
   - extends Bennedsen, Hillebrand, and Nielsen (2024)
3. Recessions and debt sustainability (Guarnieri, in progress)
   - possibility of asymmetric responses to shocks depending on sign of growth g
   - sustainability of government debt depends on sign of $g - r$
   - suggests three regimes, delimited by

$$g \leq \min\{r, 0\} \qquad \min\{r, 0\} \leq g \leq \max\{r, 0\} \qquad g \geq \max\{r, 0\}$$

Conclusion

- We have developed a flexible class of nonlinear SVARs

$$f_0(z_t) = c + \sum_{i=1}^{k} f_i(z_{t-i}) + \varepsilon_t$$

- Extended the GJRT to this setting, allowing for nonlinear common trends and attractor spaces
- Obtained long-run identifying conditions analogous to those available in the linear cointegrated SVAR
- Results provide the basis for limit distribution theory for $n^{-1/2} z_{\lfloor n\lambda \rfloor}$, and thence for inference
  - see Duffy and Jiao (2026) for some work in this direction

Bibliography

Aruoba, S. B., M. Mlikota, F. Schorfheide, and S. Villalvazo (2022): "SVARs with occasionally-binding constraints," *Journal of Econometrics*, 231, 477–499.

Ascari, G., P. Bonomolo, and Q. Haque (2023): "The long-run Phillips curve is ... a curve," De Nederlandsche Bank Working Paper No. 789.

Auerbach, A. J. and Y. Gorodnichenko (2012): "Measuring the output responses to fiscal policy," *American Economic Journal: Economic Policy*, 4, 1–27.

Beare, B. K., J. Seo, and W.-K. Seo (2017): "Cointegrated linear processes in Hilbert space," *Journal of Time Series Analysis*, 38, 1010–1027.

Beaudry, P., C. Hou, and F. Portier (2025): "On the fragility of the nonlinear Phillips curve view of recent inflation," National Bureau of Economic Research, Working Paper 33522.

Benigno, P. and G. B. Eggertsson (2023): "It's baaack: The surge in inflation in the 2020s and the return of the non-linear Phillips curve," National Bureau of Economic Research, Working Paper 31197.

Bennedsen, M., E. Hillebrand, and M. Ø. Nielsen (2024): "The Global Carbon Budget as a cointegrated system," arXiv:2412.09226.

Berry, S. T. and P. A. Haile (2018): "Identification of nonparametric simultaneous equations models with a residual index structure," *Econometrica*, 86, 289–315.

Bibliography

Blanchard, O. J. and D. Quah (1989): "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655–673.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021): "Feedbacks: financial markets and economic activity," *American Economic Review*, 111, 1845–1879.

Caggiano, G., E. Castelnuovo, V. Colombo, and G. Nodari (2015): "Estimating fiscal multipliers: News from a non-linear world," *Economic Journal*, 125, 746–776.

Duffy, J. A. and X. Jiao (2026): "Inference on common trends in a cointegrated nonlinear SVAR," *Oxford Bulletin of Economics & Statistics*, 88, 658–676.

Duffy, J. A. and S. Mavroeidis (2024): "Common trends and long-run identification in nonlinear structural VARs," arXiv:2404.05349.

——— (2026): "Identification in (endogenously) nonlinear SVARs is easier than you think," arXiv:2604.07718.

Duffy, J. A., S. Mavroeidis, and S. Wycherley (2023): "Stationarity with Occasionally Binding Constraints," arXiv:2307.06190.

——— (2025): "Cointegration with occasionally binding constraints," *Journal of Econometrics*, 252, 106103.

Engle, R. F. and C. W. J. Granger (1987): "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, 55, 251–276.

Bibliography

Gouriéroux, C., J. J. Laffont, and A. Monfort (1980): "Coherency conditions in simultaneous linear equation models with endogenous switching regimes," *Econometrica*, 48, 675–695.

Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 357–384.

Hansen, P. R. (2005): "Granger's representation theorem: a closed-form expression for $I(1)$ processes," *Econometrics Journal*, 8, 23–38.

Hubrich, K. and T. Teräsvirta (2013): "Thresholds and smooth transitions in vector autoregressive models," in *VAR Models in Macroeconomics – New Developments and Applications: essays in honor of Christopher A. Sims*.

Hubrich, K. and D. F. Waggoner (2022): "The transmission of financial shocks and leverage of financial institutions: an endogenous regime switching framework," FRB Atlanta Working Paper 2022-5.

Johansen, S. (1995): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, O.U.P.

Kilian, L. and H. Lütkepohl (2017): *Structural Vector Autoregressive Analysis*, C.U.P.

King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1991): "Stochastic Trends and Economic Fluctuations," *American Economic Review*, 81, 819–840.

Bibliography

Koop, G., M. H. Pesaran, and S. M. Potter (1996): "Impulse response analysis in nonlinear multivariate models," *Journal of Econometrics*, 74, 119–147.

Liebscher, E. (2005): "Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes," *Journal of Time Series Analysis*, 26, 669–689.

Matzkin, R. L. (2008): "Identification in nonparametric simultaneous equations models," *Econometrica*, 76, 945–978.

Mavroeidis, S. (2021): "Identification at the zero lower bound," *Econometrica*, 89, 2855–2885.

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for linear processes," *Annals of Statistics*, 20, 971–1001.

Ramey, V. A. and S. Zubairy (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850–901.

Saikkonen, P. (2008): "Stability of regime switching error correction models under linear cointegration," *Econometric Theory*, 24, 294–318.

Stock, J. H. and M. W. Watson (1988): "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

Bibliography

Teräsvirta, T., D. Tjøstheim, and C. W. J. Granger (2010): *Modelling Nonlinear Economic Time Series*, O.U.P.

Tjøstheim, D. (2020): "Some notes on nonlinear cointegration: a partial review with some novel perspectives," *Econometric Reviews*, 39, 655–673.

Tong, H. (1990): *Non-linear Time Series: a dynamical system approach*, O.U.P.