A Graphical Causal-Search Methodology for Cointegrated Systems: with an Application to Macroeconomic Variables

Kevin D. Hoover
Duke University

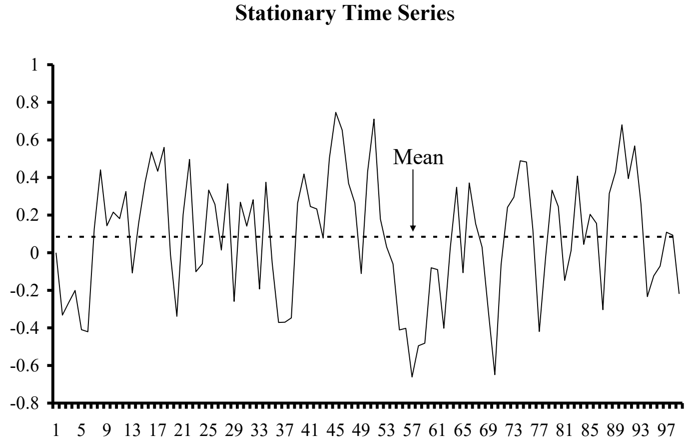
Katarina Juselius
University of Copenhagen

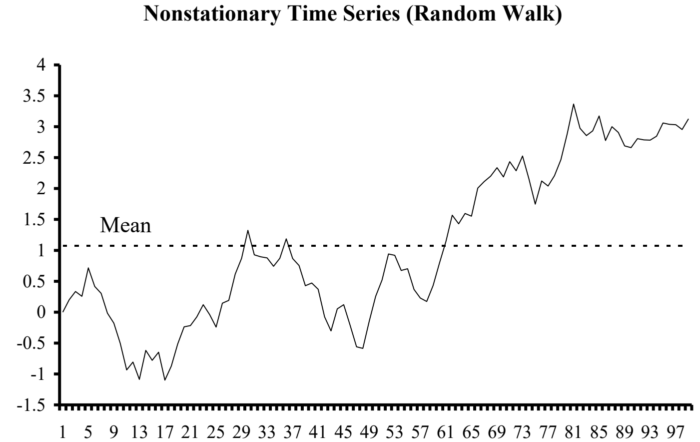
Soren Jøhansen
University of Copenhagen

Workshop on *Causal Discovery: Historical Development and State of the Art*

Carnegie Mellon University, 4 August 2023.

Stationary versus Nonstationary Time Series

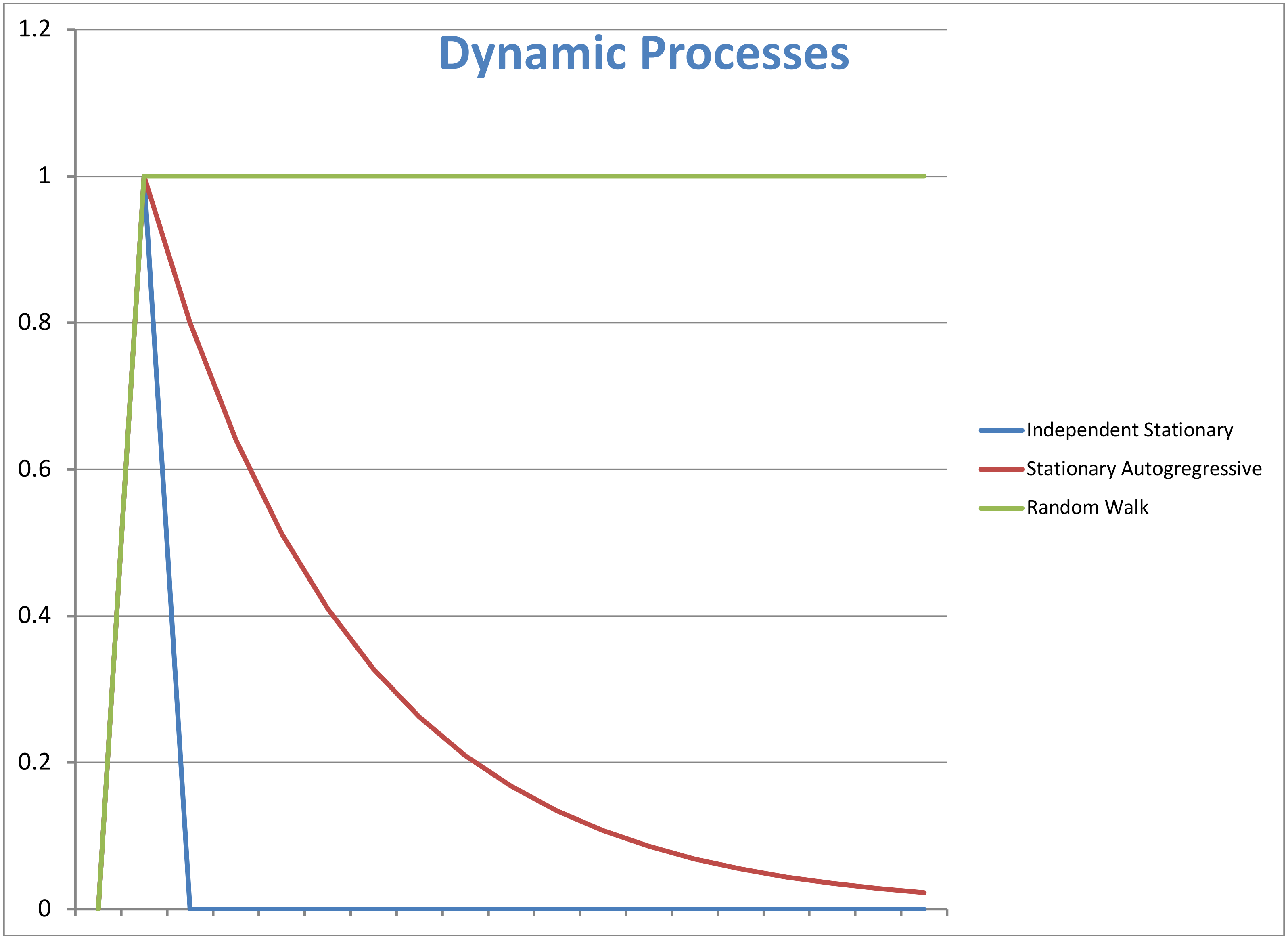




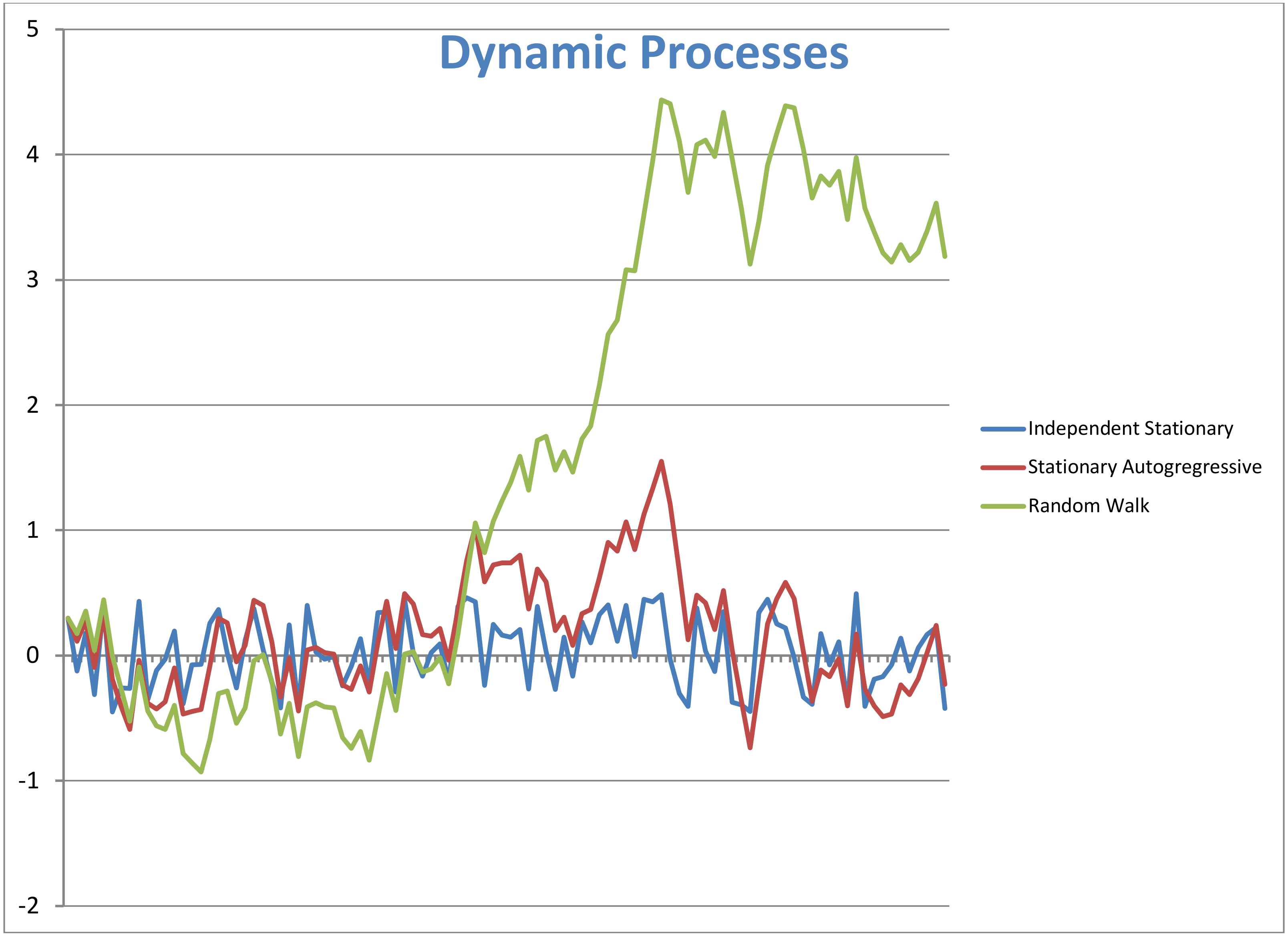
- Stationary:
    - crosses sample mean infrequently
    - mean approximately the same for all subsamples
    - *integrated of degree zero – I(0)* = need not be differenced to achieve stationarity
- Nonstationary:
    - crosses sample mean infrequently
    - mean different for different subsamples
    - *integrated of degree n – I(n)* = need not be differenced to achieve stationarity
    - I(1) series– must be differenced once to become stationary

Dynamics

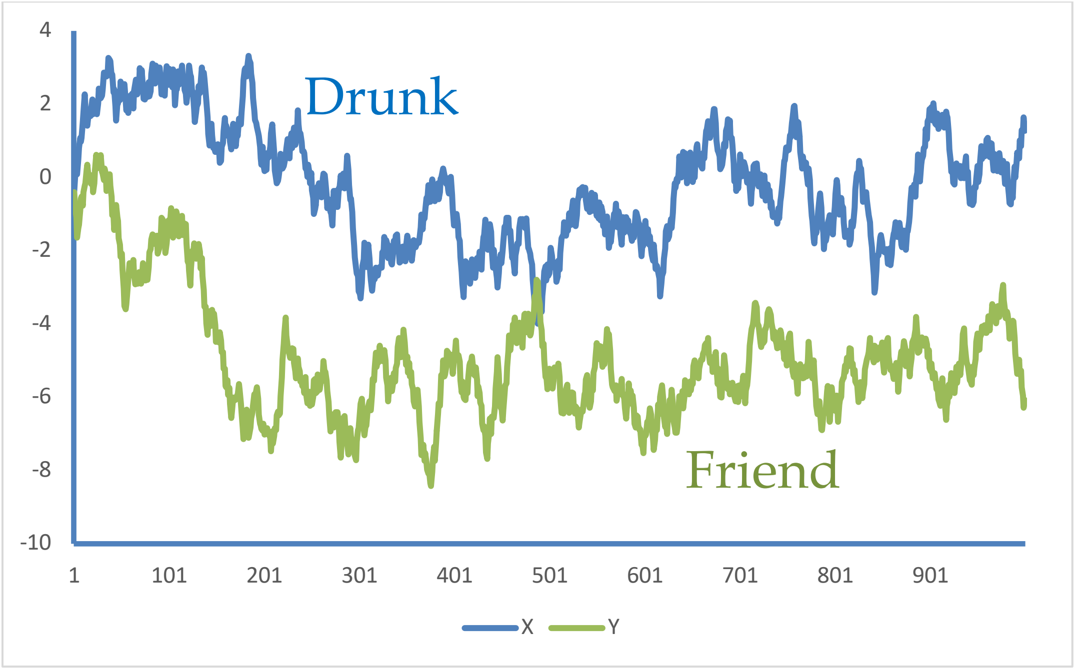
Process		Autoregressive	Moving Average
Stationary [I(0)] Independent	$\rho = 0$	$x_t = \varepsilon_t$	$x_t = \varepsilon_t$
Stationary [I(0)] Autoregressive	$\lvert\rho\rvert \underline{<} 1$	$x_t = \rho x_{t-1} + \varepsilon_t$	$x_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$
Nonstationary [I(1)] Random Walk	$\lvert\rho\rvert = 1$	$x_t = x_{t-1} + \varepsilon_t$ $\Delta x_t = x_t - x_{t-1} = \varepsilon_t$	$x_t = x_{t-n} + \sum_{j=0}^{n} \varepsilon_{t-j}$



Dynamic Processes
1.2
1
0.8
0.6
0.4
0.2
0
Independent Stationary
Stationary Autogregressive
Random Walk

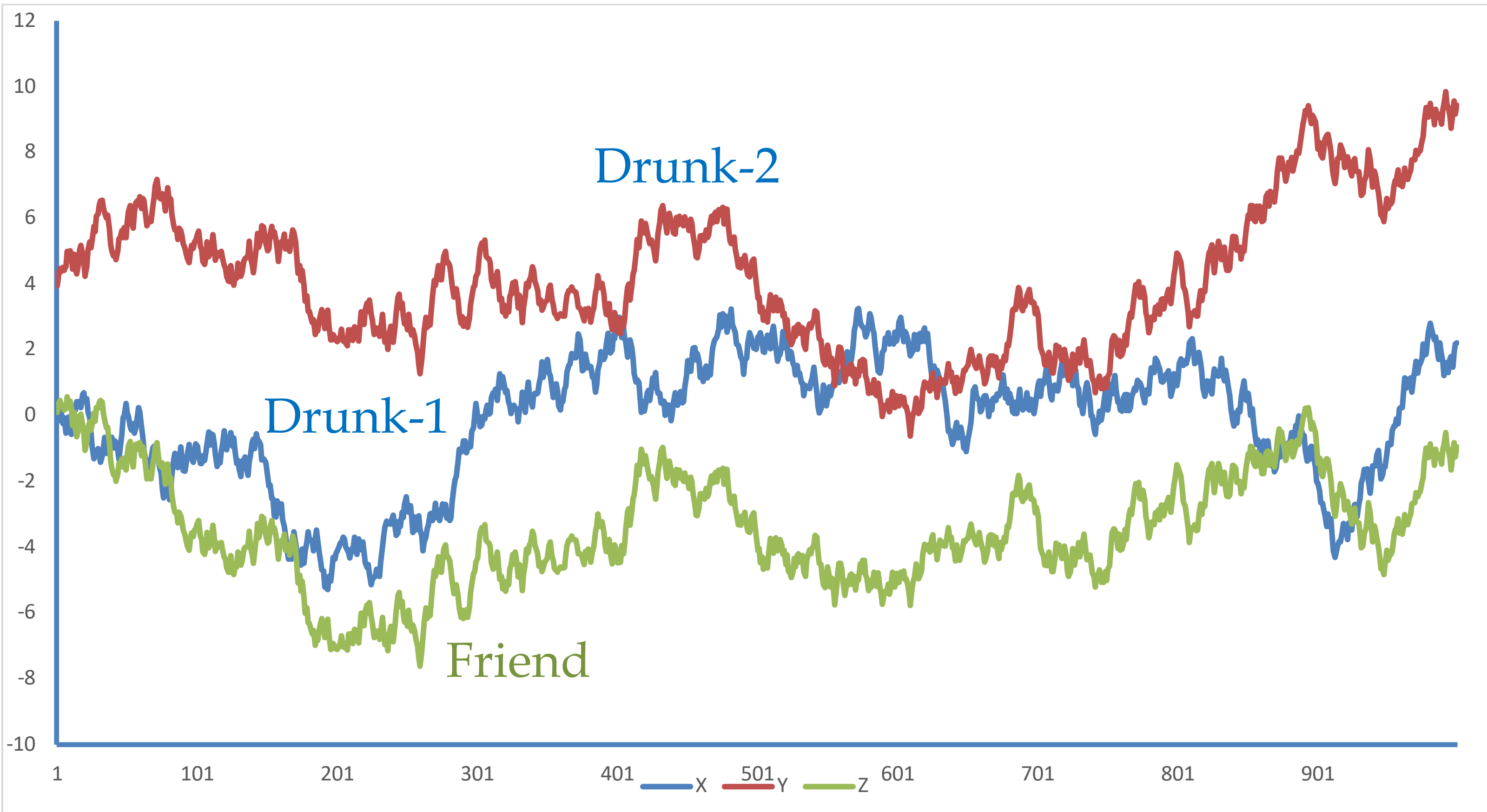

Dynamic Processes
5
4
3
2
1
0
-1
-2
Independent Stationary
Stationary Autogregressive
Random Walk

Cointegration: A Drunk and a Friend

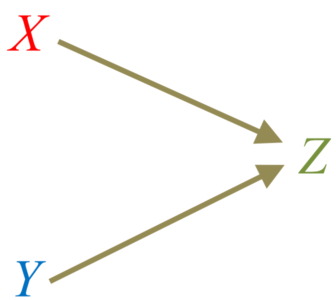


- series are I(1):
  - means: Drunk: 1st half 0.06; 2nd half –0.42;
    Friend: 1st half –4.54; 2nd half –5.34
- series *cointegrated* if individually I(1) and difference is I(0)
  - (Drunk – Friend) ~ I(0)
  - mean: 1st half –0.07; 2nd half –0.06

Cointegrated System:
Two Drunks and a Friend – 1



Cointegrated System:
Two Drunks and a Friend – 2



$\Delta X = \varepsilon_X$

[trend: $X \sim$ I(1)]

$\Delta Y = \varepsilon_Y$

[trend: $Y \sim$ I(1)]

$\Delta Z = -\delta(Z_{-1} - \theta X_{-1} - \phi Y_{-1}) + \varepsilon_Z$

(cointegrating relation:

[$(Z - \theta X - \phi Y) \sim$ I(0)

- X and Y are long-run causes of Z
- q I(1) nonstationary trends correspond to unit roots (|eigenvalues| = 1) of the companion matrix of the difference equations
- $r = p - q$ I(0) stationary cointegrating relations, where p = # of variables

Practical Metaphysical Framework

- Data-generating Process (DGP)
  - system of *ordinary variables* – i.e., variables that would be I(0) if driven only by latent I(0) independently distributed random *shocks*
  - but are I(1) if driven by (possibly *latent*) I(1) *trends*:
- 3 levels of causal relations among
  - contemporaneous variables (Y_t)
  - lagged I(0) variables (ΔY_{t-1})
  - long-run I(1) variables (Y_{t}^{∞})
- levels → interdependent dynamics; yet *nearly decomposable* (per Herbert Simon)
  - causal intervention independently possible at each level
  - causal order independent at each level (distinct causal graphs)

Structural Cointegrating Vector Autoregression (SCVAR)

$\mathbf{A\Delta Y} = \mathbf{\Gamma\Delta Y}_{-1} + \mathbf{\Phi Y}_{-1} + \boldsymbol{\varepsilon}$

contemporaneous | I(0) dynamics | long-run dynamics | shocks

$$\mathbf{A\Delta Y}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ * & 1 & 0 & 0 & 0 \\ * & * & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta X_t \\ \Delta Y_t \\ \Delta Z_t \\ \Delta T_{1,t} \\ \Delta T_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & * & * \\ 0 & 1 & 0 & * & * \\ * & * & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \\ T_{1,t-1} \\ T_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{X,t} \\ \varepsilon_{Y,t} \\ \varepsilon_{Z,t} \\ \varepsilon_{T1,t} \\ \varepsilon_{T2,t} \end{bmatrix} = \mathbf{\Phi Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- possible to estimate DGP only if trends are observed; more often estimate only using the observable variables, where fundamental trends are latent
- concentrate I(0) dynamics out of likelihood function
- must determine causal structure of contemporaneous variables to concentrate out **A**
- rank of $\mathbf{\Phi}$ = # of cointegrating relations = 3 → 2 trends
- identification requires sufficient zero restrictions to guarantee uniqueness of estimates → role of causal search

The Long-run Identification Problem – 1

$$\mathbf{A}\Delta\mathbf{Y}_t = \mathbf{\Phi}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} 1 & * & * & * & * \\ * & 1 & * & * & * \\ * & * & 1 & * & * \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \\ T_{1,t-1} \\ T_{2,t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

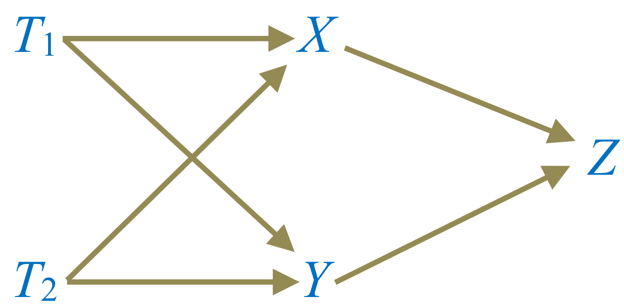
- $\mathbf{\Phi}$ is singular → reduced-rank regression, using iterative procedure: $\mathbf{a}\boldsymbol{\beta}' = \mathbf{\Phi}$
- can't apply search involving standard independence tests to $\mathbf{\Phi}$: long-run covariance matrix is singular

The Long-run Identification Problem – 2

$$\mathbf{A}\Delta\mathbf{Y}_t = \mathbf{\Phi}\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} 1 & * & * & * & * \\ * & 1 & * & * & * \\ * & * & 1 & * & * \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \\ T_{1,t-1} \\ T_{2,t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- decomposition is not unique
  - $\boldsymbol{\beta}'$ define a vector space – determines where the system is relative to long-run equilibrium
  - **a** translates deviations into adjustments towards long-run equilibrium
  - rotations of (different linear combinations of the rows of $\mathbf{a}\boldsymbol{\beta}'$ that have the same likelihood: $\mathbf{a}\boldsymbol{\beta}' = \mathbf{\Phi} = (\mathbf{a}\mathbf{Q})(\mathbf{Q}^{-1}\boldsymbol{\beta}') = \mathbf{a}^*\boldsymbol{\beta}^{*\prime}$, $\mathbf{Q}$ = full-rank conformable matrix
  - zero rows in **a** are *invariant* under rotations
- need causal search method restricting allowable rotations – possibly to a unique graph

Causal Graph and Structural Model



$$\mathbf{A}\Delta\mathbf{Y}_t = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t = \begin{bmatrix} * & 0 & 0 \\ 0 & * & 0 \\ 0 & 0 & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & * & * \\ 0 & 1 & 0 & * & * \\ * & * & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \\ T_{1,t-1} \\ T_{2,t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t$$

Basis of Causal Search – 1

- *Weak Exogeneity*:
  - set of variables **X** is *weakly exogenous* for a set **Y** iff joint probability distribution of can be decomposed:

    conditional distribution of Y|X × *marginal distribution of X*,
    where their parameters are mutually unconstraining.
  - required for efficient estimation of conditional distribution
  - parameterization-relative; thus holding in a set does not imply holding in a subset
  - *long-run* weak exogeneity corresponds to a zero row in **a** → testable
    - recall: zero rows in **a** are invariant under rotations of $\mathbf{a}\boldsymbol{\beta}' = \boldsymbol{\Phi}$
- *Johansen's Lemma*:

$$\mathbf{a} = \boldsymbol{\Sigma}(\mathbf{M}_{12}\mathbf{V}_{2\mathbf{T}} + \mathbf{C}_1\mathbf{V}_{\mathbf{TT}})_{\perp}$$

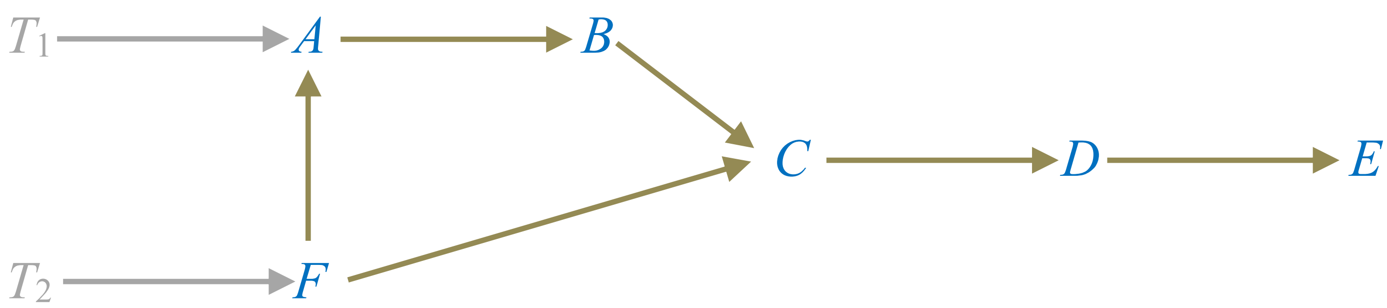
  - given the *causally-ordered* DGP, formula allows calculation the implied weak-exogeneity relations (i.e., zero rows in **a**)
  - *cointegration* is easily calculated, given the causal structure

Basis of Causal Search – 2

- *Irreducible Cointegration*:
  - a set of variables is *irreducibly cointegrated* if it is cointegrated and it contains no smaller cointegrated sets
  - CVARs of all irreducibly cointegrated sets have rank = 1 → testable

- *Main Theorem*:
  - *for any irreducibly cointegrated set of p variables, if p – 1 variables are weakly exogenous for the p*th *variable, then the weakly exogenous variables form an* observed collider *at the p*th *variable*
    - proof uses Johansen's lemma
    - examples of observed collider:

Examples of Observed Colliders



Pairs

- $\{A, \boldsymbol{B}\}$
- $\{C, \boldsymbol{D}\}$
- $\{C, \boldsymbol{E}\}$
- $\{D, \boldsymbol{E}\}$

Triples

- $\{A, \boldsymbol{C}, F\}$
- $\{A, \boldsymbol{D}\ F\}$
- $\{A, \boldsymbol{E}, F\}$
- $\{B, \boldsymbol{C}, F\}$
- $\{B, \boldsymbol{D}, F\}$
- $\{B, \boldsymbol{E}, F\}$

- largest possible IC set: # *of trends* + 1 = 2 +1 = 3
- colliders create *local* trends – i.e., linear combinations of (usually unobserved) *fundamental* trends
- shielded unshielded collider distinction not important
- any unobserved intermediate variable must share local trend with observed parent

Search Procedure Illustrated

Step 1: Initial Specification

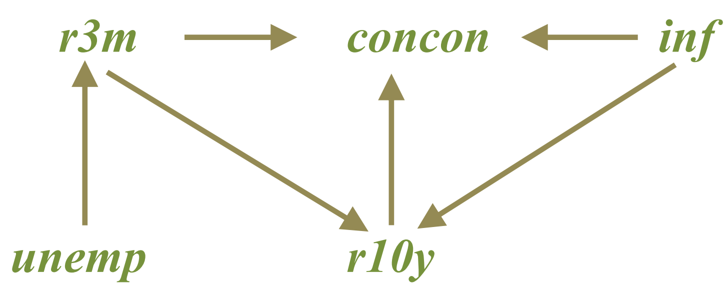
- I(1) variables:
  - *concon*: Consumer Confidence
  - *inf*: CPI inflation
  - *r3m*: yield on 3-month Treasury bill
  - *r10y*: yield on 10-year Treasury bond
  - *unemp*: unemployment rate
- time period:
  - monthly data: 1985: June – 2007: June
  - after monetary policy upheavals of the early 1980s; before the financial crises of 2007-2009
- other details:
  - 3 lags of levels, allows 2 lags of 1st differences
  - constant restricted to cointegrating relations
  - dummy variables to control for outliers

Step 2: System Rank Determination

- *rank* (*r*) = 3 based on standard Johansen test for system cointegration
- implies # *of cointegrating relations* = 3; # *of trends* (*q*)= 2

Step 3: Specify Contemporaneous Causal Order

- uses PC algorithm specialized to the SVAR based on earlier work
  - Swanson & Granger (1997)
  - Demiralp & Hoover (2003)
  - Demiralp, Hoover, & Perez (2008)
- selected graph



Step 4: Estimate SCVAR

$$\mathbf{A}\Delta\mathbf{Y}_t = \begin{bmatrix} 1 & * & * & * & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ 0 & * & * & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta concon_t \\ \Delta inf_t \\ \Delta r3m_t \\ \Delta r10y_t \\ \Delta unemp_t \end{bmatrix} = \mathbf{a}\boldsymbol{\beta}' \begin{bmatrix} concon_t \\ inf_t \\ r3m_t \\ r10y_t \\ unemp_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{X,t} \\ \varepsilon_{Y,t} \\ \varepsilon_{Z,t} \\ \varepsilon_{T1,t} \\ \varepsilon_{T2,t} \end{bmatrix} = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- encodes contemporaneous causal graph in **A**

Step 5: Concentrate Out Contemporaneous Causal Relations

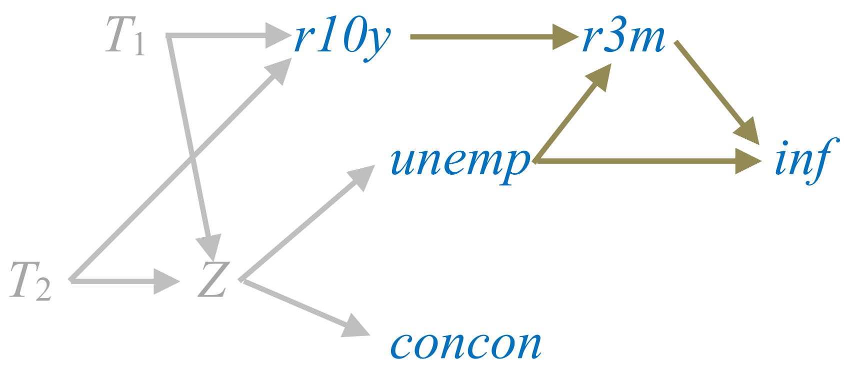
$$\Delta \mathbf{Y}_t = \begin{bmatrix} \Delta concon_t \\ \Delta inf_t \\ \Delta r3m_t \\ \Delta r10y_t \\ \Delta unemp_t \end{bmatrix} = \mathbf{a\beta}' \begin{bmatrix} concon_t \\ inf_t \\ r3m_t \\ r10y_t \\ unemp_t \end{bmatrix} + \begin{bmatrix} \varepsilon_{X,t} \\ \varepsilon_{Y,t} \\ \varepsilon_{Z,t} \\ \varepsilon_{T1,t} \\ \varepsilon_{T2,t} \end{bmatrix} = \mathbf{a\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- note the abuse of notation
  - **Y** is not the same as when **A** is present
  - $\mathbf{a\beta}'$ and $\boldsymbol{\varepsilon}$ are identical
  - long-run information is identical
- long-run not yet causally ordered
  - just one possibility within the vector space spanned by $\boldsymbol{\beta}$,such that $\mathbf{a\beta}' = \boldsymbol{\Phi}$
  - $\mathbf{a}$ conforms to non-unique $\boldsymbol{\beta}'$

Step 6: Identify IC Sets & Weak Exogenity

- *irreducible cointegration*: find all subsets of the variables ≤ # of trends + 1 and test for cointegration rank = 1
  - 10 possible 2-member subsets
  - 10 possible 3-member subsets
- *weak exogeneity*: test those subsets for zero rows in **a**
- *ignoring ambiguous cases* →
  - 1 noncolliding IC set:
    - {*concon*, *unemp*}
  - 3 observed colliders:
    - {*r10y*, *unemp*} $\mapsto$ *r3m*
    - {*r10y*, *unemp*} $\mapsto$ *inf*
    - {*r3m*, *unemp*} $\mapsto$ *inf*
  - an unsettled anomaly with *concon* is ignored for the present – *needs further work*

Step 7: Identify Causal Graph Consistent with Observed Colliders



♦

$$\Delta \mathbf{Y}_t = \begin{bmatrix} \Delta concon_t \\ \Delta inf_t \\ \Delta r3m_t \\ \Delta r10y_t \\ \Delta unemp_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & * \\ * & 0 & 0 \\ 0 & * & 0 \\ \Diamond & \Diamond & \Diamond \\ 0 & 0 & * \end{bmatrix} \begin{bmatrix} 0 & 1 & * & 0 & * \\ 0 & 0 & 1 & * & * \\ * & \Diamond & \Diamond & \Diamond & 1 \end{bmatrix} \begin{bmatrix} concon_t \\ inf_t \\ r3m_t \\ r10y_t \\ unemp_t \end{bmatrix} + \boldsymbol{\varepsilon}_t = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- Test of Overidentifying Restrictions: $\chi^2(6) = 7.4$, $p = 0.48$

Step 8: Sharpen the Representation & Estimates

- general-to-specific search to eliminate insignificant not-casually ordered regressors

$$\Delta \mathbf{Y}_t = \begin{bmatrix} \Delta concon_t \\ \Delta inf_t \\ \Delta r3m_t \\ \Delta r10y_t \\ \Delta unemp_t \end{bmatrix} = \begin{bmatrix} 0 & 0 & * \\ * & 0 & 0 \\ 0 & * & 0 \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix} \begin{bmatrix} 0 & 1 & * & 0 & * \\ 0 & 0 & 1 & * & * \\ * & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} concon_t \\ inf_t \\ r3m_t \\ r10y_t \\ unemp_t \end{bmatrix} + \boldsymbol{\varepsilon}_t = \mathbf{a}\boldsymbol{\beta}'\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t$$

- Test of Overidentifying Restrictions: $\chi^2(11) = 8.3$, $p = 0.83$

Thanks

The End