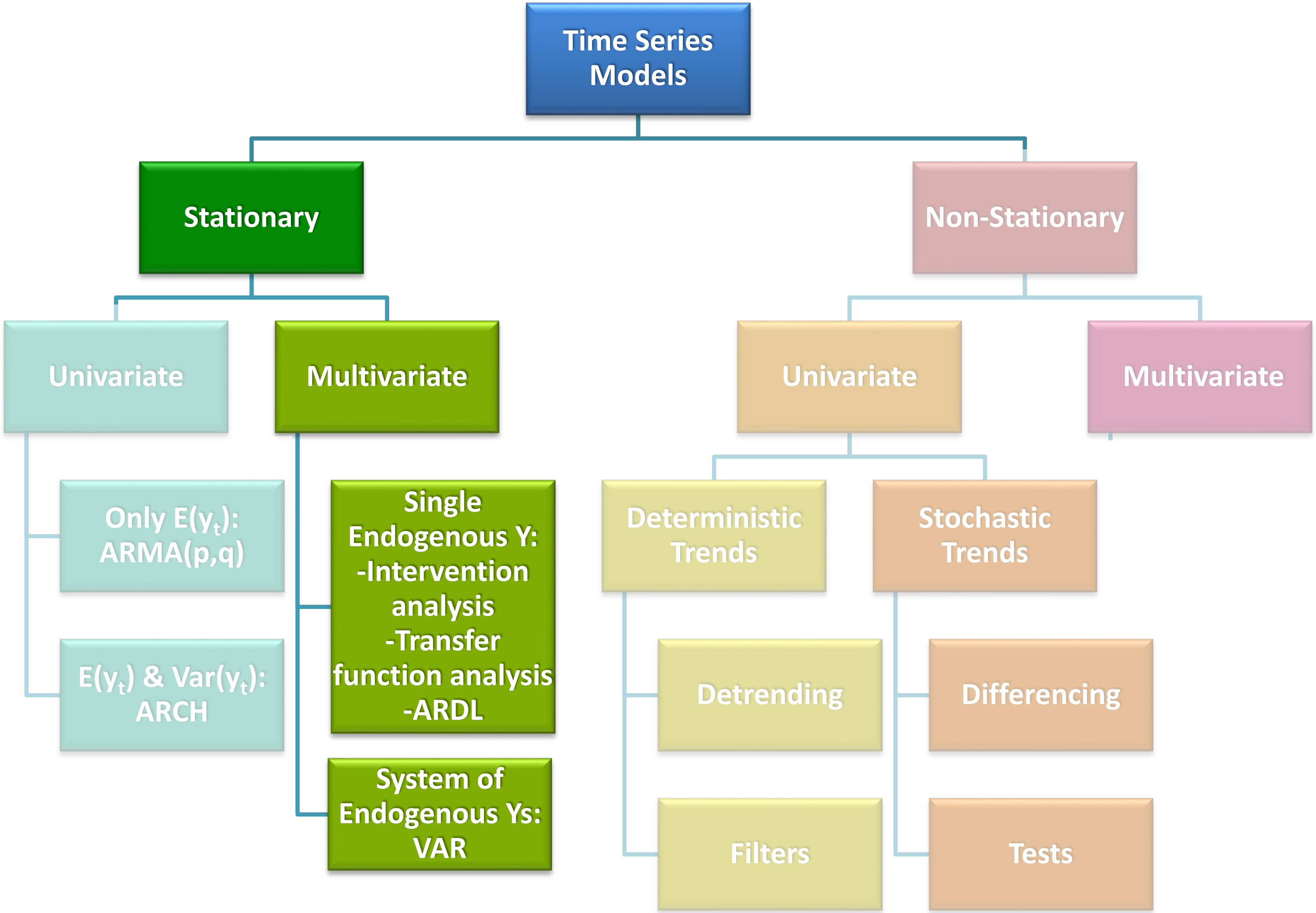
Advanced Time Series Econometrics

TOPIC 3:

Stationary Multivariate Models


Time Series Models
Stationary
Non-Stationary
Univariate
Multivariate
Univariate
Multivariate
Only $E(y_t)$: ARMA(p,q)
$E(y_t)$ & $Var(y_t)$: ARCH
Single Endogenous Y: -Intervention analysis -Transfer function analysis -ARDL
System of Endogenous Ys: VAR
Deterministic Trends
Stochastic Trends
Detrending
Filters
Differencing
Tests

motivation of sequencing

- Note that I follow a different sequence in the course than Enders does in the book
  - I do the stationary multivariate models first as they are a natural extension of the univariate
  - This forms part of the mathematical toolkit for a time series econometrician.
  - There is broad consensus on these parts.
  - The non-stationary part econometrics looks very different, and is much more diverse.
  - Thus, we build the full toolkit assuming stationarity first, and then use it to understand the complications and changes brought if processes can be non-stationary

Plan

- Introduction
- We will briefly study:
  - Intervention analysis
  - Transfer function Analysis
  - Autoregressive, distributed lag models (ADL,ARDL)
    - We will return to this approach in depth later.
- Then in depth:
  - Vector Autoregression
    - Structural vs Reduced form eq'ns: Identification
    - Stationarity
    - Analyzing the information in a VAR
    - Back to Identification – Various approaches
    - Estimation Methods
- Along the way we will do a joyful review of your love of linear algebra results

Standard Macro Model

- Goal is to estimate standard macroeconomic models like the typical 3 equation New Keynesian model.

Phillips' Curve: $$\pi_t = \pi_{t-1} + \lambda y_t - \gamma r_t + \varepsilon_{\pi,t}$$

IS Curve: $$y_t = E_t\left[y_{t+1}\right] - \theta r_t + \varepsilon_{y,t}$$

Monetary Policy rule: $$r_t = \phi_\pi \pi_t + \phi_y y_t + \varepsilon_{r,t}$$

Problem:

- If we were to try to estimate these equations individually,
  - OLS will give inconsistent estimates
  - RHS variables *endogenous = correlated with error term*

$$
\begin{aligned}
\text{Phillips' Curve: } \pi_t &= \pi_{t-1} + \lambda y_t - \gamma r_t + \varepsilon_{\pi,t} \\
\text{IS Curve: } y_t &= E_t\left[y_{t+1}\right] - \theta r_t + \varepsilon_{y,t} \\
\text{Monetary Policy rule: } r_t &= \phi_\pi \pi_t + \phi_y y_t + \varepsilon_{r,t}
\end{aligned}
$$

Multivariate models

- I will present a sequence of models where the new variables we introduce range from fully exogenous to mutually endogenous
- Using the familiar OLS assumptions as an organizing tool
- We start with an endogenous variable of interest, which last week we modelled as a univariate process and add additional right hand side variables as explanatory features

OLS Assumptions

$$y = X\beta + \varepsilon$$

1. Linearity
2. No perfect multicollinearity
3. "Exogeneity of the X's"
4. Zero mean, unsystematic errors
5. Normal errors

OLS Assumptions

1. Linearity
2. Full column rank X
3. “Exogeneity of the X’s”

- Different assumptions required for the β estimate to be
    - Unbiased, or
    - Consistent

Exogeneity: from strong to weak

- X is non-stochastic
- Mean independence
- “Same period” mean independence
- Predetermined variables on RHS

Intervention Analysis

- Formal test of the impact of a non-stochastic event or change on the mean of a time series:

$$y_t = a_0 + a_1 y_{t-1} + c_0 z_t + \varepsilon_t$$

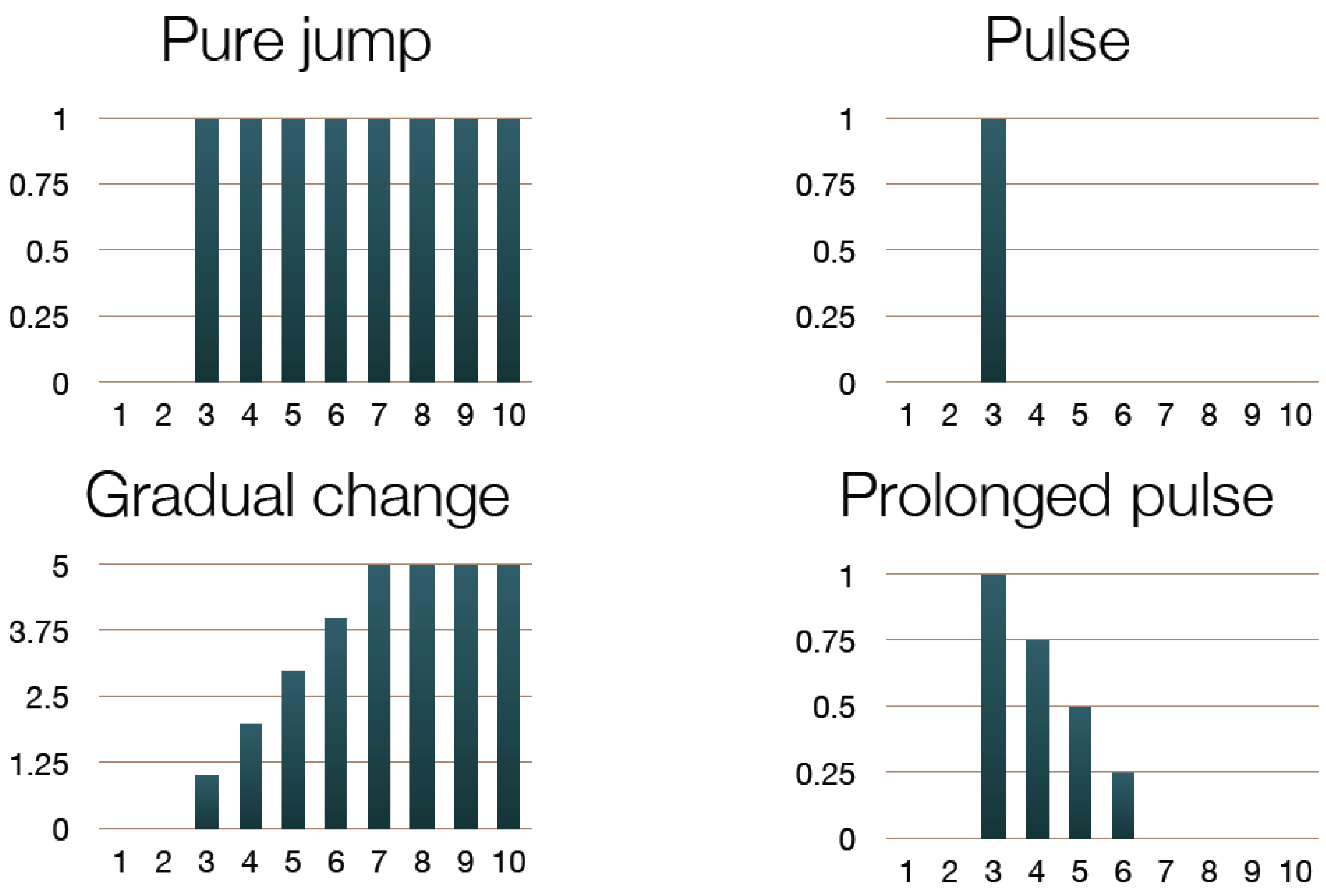
  - z_t represents an "impulse" that causes change
  - Modelled as a dummy variable – non stochastic

- Enders' example:
  - Impact of metal detectors on sky-jackings

- Impact, path and long run effects can be characterized
- One can also allow the impact to change the coefficients of the process
  - I.e. that induces a structural change

Some possible Impulse Patterns

Disappearance of MH350 on demand for Malaysian Airlines flights?



Problems with Intervention Analysis

Table 5.1 Metal Detectors and Skyjackings

	Pre-Intervention Mean	a_1	Impact Effect (c_0)	Long-Run Effect
Transnational $\{TS_t\}$	3.032	0.276	−1.29	−1.78
	(5.96)	(2.51)	(−2.21)	
US domestic $\{DS_t\}$	6.70		−5.62	−5.62
	(12.02)		(−8.73)	
Other skyjackings $\{OS_t\}$	6.80	0.237	−3.90	−5.11
	(7.93)	(2.14)	(−3.95)	

Notes:

[1] *t*-Statistics are in parentheses.

[2] The long-run effect is calculated as $c_0/(1 - a_1)$.

- Enders gives an excellent example that is detailed
  - What is required for the method to be reliable?
  - What are the limitations?
- Lucas-critique
  - Solution?
- Limited applications – requires change to be fully exogenous

Exogeneity: from strong to weak

- X is non-stochastic
- Mean independence

$$E(z_t \varepsilon_s) = 0 \ \forall \ t, s$$

- "Same period" mean independence
- Predetermined variables on RHS

Transfer function and ARDL models

- Impact of a stochastic, exogenous **process** on dependent variable
- Impact impulse dynamic (i.e. itself an interesting ARMA(p,q) process
- And the variable of interest may itself have ARMA(p’,q’) properties
- Goal is to find a parsimonious model of the (ONE WAY) interaction

Note on the treatment in Enders

- New in this version
- Enders discusses this as a general ARDL but the application and derivations all clearly assume strictly exogenous additional variables
- How ARDL models are currently employed is not like this. Typically mutually endogenous variables are modelled as a "single equation ARDL", and endogeneity is dealt with in some other way (Instrumental Variables etc.)

Transfer function and ARDL models

- Impact of a stochastic, exogenous process on dependent variable
  - E.g. impact of **eruptions of Eyjafjallajökull** on airline revenue/profits, tourism, farming output in Iceland
  - Enders' example: impact of terrorism on tourism
- Impact may be dynamic, complicated
  - Patterns identified via **cross-correlogram**
- See the excellent example in Enders and the caveats

Exogeneity: from strong to weak

- X is non-stochastic
- Mean independence
- "Same period" mean independence

$$E(z_t \varepsilon_s) = 0 \; \forall \; s \geq t$$
$$E(z_t \varepsilon_s) \neq 0 \; \forall \; s < t$$

- Predetermined variables on RHS

But

- Standard macro most interested in jointly endogenous processes:

Phillips' Curve: $\pi_t = \pi_{t-1} + \lambda y_t - \gamma r_t + \varepsilon_{\pi,t}$

IS Curve: $y_t = E_t\left[y_{t+1}\right] - \theta r_t + \varepsilon_{y,t}$

Monetary Policy rule: $r_t = \phi_\pi \pi_t + \phi_y y_t + \varepsilon_{r,t}$

Exogeneity: from strong to weak

- X is non-stochastic
- Mean independence
- “Same period” mean independence
- The most common situation in macroeconomics is a mutually endogenous set of variables
  - The strongest assumption we can reasonably make is that variables are **predetermined:**
  - The errors of the process are pure *innovations*:

$$E(z_t \varepsilon_s) = 0 \ \forall \ s > t$$
$$E(z_t \varepsilon_s) \neq 0 \ \forall \ s \leq t$$

History of Empirical Macro (pre 1980)

- Large structural economic models, loosely informed by theory
  - Estimate each equation individually
  - Aggregate results, forecast
- The process followed implied that the following is known with certainty:
  - Relationships between variables (which variables affects which ones in which sequence)
  - Lag structure of DGP

The contribution of Sims (1980)

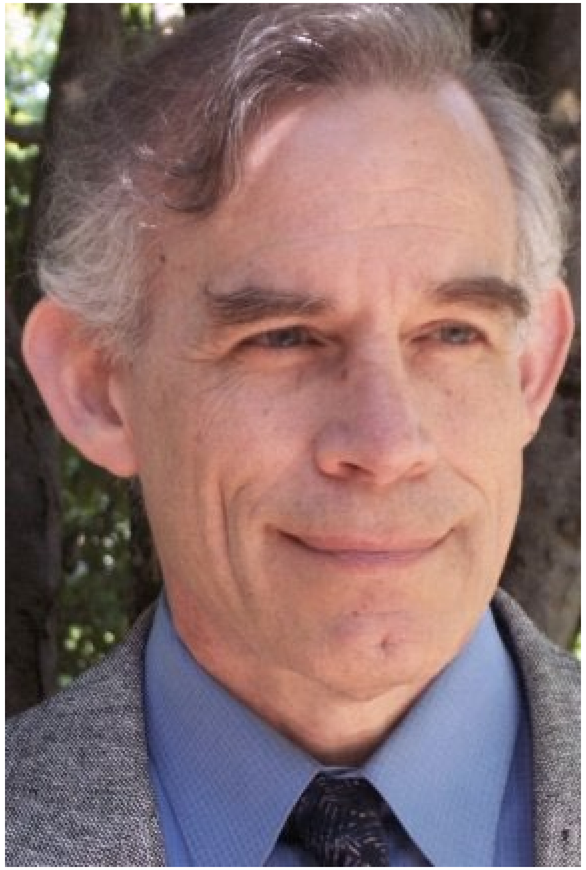
- Chris Sims publishes his seminal criticism of empirical macro models in 1980
- Key concern: "Incredible Restrictions"
  - Relevant variables
  - Precise timing of feedback

Brookings Quarterly Econometric Model of the United States, as reported by Suits and Sparks (p. 208, 1965):

$$C_{\mathrm{NF}} = \underset{(0.0165)}{0.0656} Y_{\mathrm{D}} - \underset{(2.49)}{10.93} (P_{\mathrm{CNF}}/P_{\mathrm{C}})_{t-1} + \underset{(0.0522)}{0.1889} (N + N_{\mathrm{ML}})_{t-1}$$

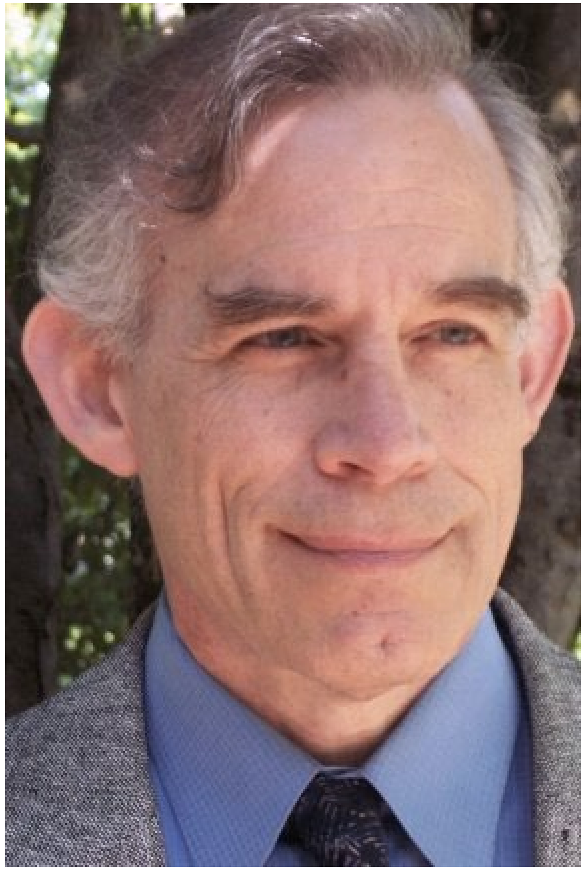
$$C_{\mathrm{NEF}} = 4.2712 + \underset{(0.0127)}{0.1691} Y_{\mathrm{D}} - \underset{(0.0213)}{0.0743} (ALQD_{\mathrm{HH}}/P_{\mathrm{C}})_{t-1}$$

where C_{NF} = personal consumption expenditures on food
Y_{D} = disposable personal income
P_{CNF} = implicit price deflator for personal consumption expenditures on food
P_{C} = implicit price deflator for personal consumption expenditures
N = civilian population
N_{ML} = military population including armed forces overseas
C_{NEF} = personal consumption expenditures for nondurables other than food
$ALQD_{\mathrm{HH}}$ = end-of-quarter stock of liquid assets held by households



Sims' contribution

- Chris Sims' publishes his seminal criticism of empirical macro models in 1980
- Key concern: "Incredible Restrictions"
  - Relevant variables
  - Precise timing of feedback
- Proposed:
  - Large estimated models with as few restrictions as possible
  - All variables endogenous, so simultaneously estimated system
- Nobel prize in 2011 for
  - "empirical research on cause and effect in the macro economy"



Plan

- Vector Autoregression
  - **Structural vs Reduced form eq’ns**
  - The Identification Problem
  - Stationarity
  - Analyzing the information in a VAR
  - Back to Identification – Various approaches
  - Estimation Methods

VAR

- Structural form of the simplest VAR model:

$$
\begin{aligned}
y_t &= \gamma_{01} + b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{y,t} \\
z_t &= \gamma_{02} + b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

Where the **structural innovations** are uncorrelated, white noise:

$E(z_{t-1}\varepsilon_{y,t}) = 0$ and $E(y_{t-1}\varepsilon_{z,t}) = 0$

$$
E\begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix} = E[\boldsymbol{\varepsilon}_t] = \mathbf{0}
$$

$$
E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'] = \begin{bmatrix} \sigma_{\varepsilon_y}^2 & 0 \\ 0 & \sigma_{\varepsilon_z}^2 \end{bmatrix}
$$

$$
E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s'] = \mathbf{0} \, \forall \, t \neq s
$$

VAR

- Structural form of the simplest model:

$$
\begin{aligned}
y_t &= \gamma_{01} + b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{y,t} \\
z_t &= \gamma_{02} + b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

- In matrix form:

$$
\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}
$$

VAR

- Structural form of the simplest model:

$$
\begin{aligned}
y_t &= \gamma_{01} + b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{y,t} \\
z_t &= \gamma_{02} + b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

- In matrix form:

$$
\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} 0 & b_{12} \\ b_{21} & 0 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}
$$

VAR

- **Structural form** of the simplest model:

$$B\mathbf{x}_t = \Gamma_0 + \Gamma_1 \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

- The **Reduced form** of the model is:

$$\begin{aligned} \mathbf{x}_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 \mathbf{x}_{t-1} + B^{-1}\boldsymbol{\varepsilon}_t \\ &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \end{aligned}$$

- where

$$\begin{aligned} \boldsymbol{e}_t &= B^{-1}\boldsymbol{\varepsilon}_t = C^{(0)}\boldsymbol{\varepsilon}_t = \begin{bmatrix} c_{11}^{(0)} & c_{12}^{(0)} \\ c_{21}^{(0)} & c_{22}^{(0)} \end{bmatrix} \boldsymbol{\varepsilon}_t \\ &= \begin{bmatrix} c_{11}^{(0)}\varepsilon_{y,t} + c_{12}^{(0)}\varepsilon_{z,t} \\ c_{21}^{(0)}\varepsilon_{y,t} + c_{22}^{(0)}\varepsilon_{z,t} \end{bmatrix} \end{aligned}$$

VAR

- **Structural form** of the simplest model:

$$B\mathbf{x}_t = \Gamma_0 + \Gamma_1 \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

- The **Reduced form** of the model is:

$$\begin{aligned} \mathbf{x}_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 \mathbf{x}_{t-1} + B^{-1}\boldsymbol{\varepsilon}_t \\ &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \end{aligned}$$

- Moreover:

$$E\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right) = E\left(C^{(0)} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' C^{(0)\prime}\right)$$

Plan

- Vector Autoregression
  - Structural vs Reduced form eq’ns
  - **The Identification Problem**
  - Stationarity
  - Analyzing the information in a VAR
  - Back to Identification – Various approaches
  - Estimation Methods

Identification Problem

- Identification
  - Recovery of Structural form from Reduced form
  - Let us count
- Structural form

$$B\mathbf{x}_t = \Gamma_0 + \Gamma_1 \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$$

$$\begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix} \quad E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right] = \begin{bmatrix} \sigma_{\varepsilon_y}^2 & 0 \\ 0 & \sigma_{\varepsilon_z}^2 \end{bmatrix}$$

- Reduced form

$$\begin{aligned} \mathbf{x}_t &= B^{-1}\Gamma_0 + B^{-1}\Gamma_1 \mathbf{x}_{t-1} + B^{-1}\boldsymbol{\varepsilon}_t \\ &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \end{aligned}$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{01} \\ a_{02} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} e_{y,t} \\ e_{z,t} \end{bmatrix} \quad E\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right) = \begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix}$$

Identification

- Identification
  - Definition: Recovery of Structural form from Reduced form
  - A key issue in the literature is being able to *identify* exogenous shocks attributable to specific policies
  - Without this, an empirical model cannot be said to have any clear implication for policy analysis
- For an unrestricted VAR, the most general structural form is under-identified by the closest estimable reduced form
  - Identification requires a restriction on the **structural** parameters
  - Note that this restriction is **entirely untestable.** It is a restriction on the *theory* not the specification
  - As such, many identification schemes have developed
  - We will consider the most basic one now, then more later

Identification

- Identification
  - Recovery of Structural form from Reduced form
- Identification Schemes (1)
  - Choleski decomposition
    - Timing assumption – which variables can contemporaneously affect which others
    - The unrestricted structural model was:

$$
\begin{aligned}
y_t &= \gamma_{01} + b_{12} z_t + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{y,t} \\
z_t &= \gamma_{02} + b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

Identification

- Identification
  - Recovery of Structural form from Reduced form
- Identification Schemes (1)
  - Choleski decomposition
    - Timing assumption – which variables can contemporaneously affect which others
    - Imposing that z cannot contemporaneously affect y:

$$
\begin{aligned}
y_t &= \gamma_{01} + \gamma_{11} y_{t-1} + \gamma_{12} z_{t-1} + \varepsilon_{y,t} \\
z_t &= \gamma_{02} + b_{21} y_t + \gamma_{21} y_{t-1} + \gamma_{22} z_{t-1} + \varepsilon_{z,t}
\end{aligned}
$$

Identification

- Identification
  - Recovery of Structural form from Reduced form
- Identification Schemes (1)
  - Choleski decomposition
    - Timing assumption – which variables can contemporaneously affect which others
    - Imposing that z cannot contemporaneously affect y:

$$\begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}$$

Identification

$$\begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}$$

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Identification

$$\begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \gamma_{01} \\ \gamma_{02} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}$$

$$\begin{aligned} B &= \begin{bmatrix} 1 & 0 \\ -b_{21} & 1 \end{bmatrix} \\ B^{-1} &= \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \boldsymbol{e}_t &= B^{-1}\boldsymbol{\varepsilon}_t = \begin{bmatrix} 1 & 0 \\ b_{21} & 1 \end{bmatrix} \boldsymbol{\varepsilon}_t \\ &= \begin{bmatrix} \varepsilon_{y,t} \\ b_{21}\varepsilon_{y,t} + \varepsilon_{z,t} \end{bmatrix} \end{aligned}$$

n Variable VAR(p)

- Primitive Form:

$$\underset{[n\times n]}{\mathbf{B}}\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\mathbf{\Gamma}_0} + \underset{[n\times n]}{\mathbf{\Gamma}_1}\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \mathbf{\Gamma}_p\mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$E\left[\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right] = \underset{[n\times n]}{\mathbf{\Sigma}_\varepsilon}$$

- Reduced Form:

$$\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\mathbf{A}_0} + \underset{[n\times n]}{\mathbf{A}_1}\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \mathbf{A}_p\mathbf{x}_{t-p} + \boldsymbol{e}_t$$

$$E\left[\mathbf{e}_t\mathbf{e}_t'\right] = \underset{[n\times n]}{\mathbf{\Sigma}}$$

n Variable VAR(p)

- Primitive Form:

$n^2 - n$ unrestricted parameters

$$\underset{[n\times n]}{\mathbf{B}}\;\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\boldsymbol{\Gamma}_0} + \underset{[n\times n]}{\boldsymbol{\Gamma}_1}\;\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \boldsymbol{\Gamma}_p \mathbf{x}_{t-p} + \boldsymbol{\varepsilon}_t$$

$$E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t'\right] = \underset{[n\times n]}{\boldsymbol{\Sigma}_{\varepsilon}} \longleftarrow n \text{ non-zero parameters}$$

- Reduced Form:

$$\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\mathbf{A}_0} + \underset{[n\times n]}{\mathbf{A}_1}\;\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \mathbf{A}_p \mathbf{x}_{t-p} + \mathbf{e}_t$$

$$E\left[\mathbf{e}_t \mathbf{e}_t'\right] = \underset{[n\times n]}{\boldsymbol{\Sigma}} \longleftarrow \frac{n(n+1)}{2} \text{ unique parameters}$$

$\frac{n^2-n}{2}$ restrictions necessary for identification

Plan

- Vector Autoregression
  - Structural vs Reduced form eq’ns
  - The Identification Problem
  - **Stationarity**
  - Analyzing the information in a VAR
  - Back to Identification – Various approaches
  - Estimation Methods

Stationarity of a VAR(1)

- We now extend our rule for stationarity of an AR(p) process to a VAR(p) process
- There are a myriad of ways of doing this, I will give a version that is not a proof but gives my favourite intuition. Enders does this another way, but it is equivalent.
- As in the univariate case, a process is stationary if the effect of shocks infinitely far in the past eventually fade out
- Again, we start with the simplest version, a 2 variable VAR(1)

$$\mathbf{x}_t \quad = \quad A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t$$

Stationarity of a VAR(1)

- As with the univariate case, we can iterate backwards:

$$\begin{aligned}
\mathbf{x}_t &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \\
&= A_0 + A_1 \left(A_0 + A_1 \mathbf{x}_{t-2} + \boldsymbol{e}_{t-1} \right) + \boldsymbol{e}_t \\
&= A_0 + A_0 A_1 + \boldsymbol{e}_t + A_1 \boldsymbol{e}_{t-1} + A_1^2 \left(A_0 + A_1 \mathbf{x}_{t-3} + \boldsymbol{e}_{t-2} \right) \\
&= A_0 + A_0 A_1 + A_0 A_1^2 + \boldsymbol{e}_t + A_1 \boldsymbol{e}_{t-1} + A_1^2 \boldsymbol{e}_{t-2} + A_1^3 \left(A_0 + A_1 \mathbf{x}_{t-4} + \boldsymbol{e}_{t-3} \right) \\
&= A_0 \sum_{i=0}^{\infty} A_1^i + \sum_{i=0}^{\infty} A_1^i \boldsymbol{e}_{t-i} + \lim_{i \to \infty} A_1^i \mathbf{x}_{t-i}
\end{aligned}$$

- To evaluate these limits, we need the tools for integer powers of square matrices.
- For this, we detour into a review of eigenvalues
- What are eigenvalues? Intuitively? *evil smiling face*

Eigenvalues

- Let
  - A be an arbitrary $[n \times n]$ matrix
  - $\mathbf{w}$ be an arbitrary, non-zero $[n \times 1]$ vector, and
  - I be the $[n \times n]$ identity matrix
- Then λ is called an *eigenvalue* or a *characteristic root* of the matrix A if:

$$A\mathbf{w} = \lambda\mathbf{w}$$

$$(A - \lambda I)\,\mathbf{w} = 0$$

Eigenvalues

- Let
  - A be an arbitrary $[n \times n]$ matrix
  - $\mathbf{w}$ be an arbitrary, non-zero $[n \times 1]$ vector, and
  - I be the $[n \times n]$ identity matrix
- Then λ is called an *eigenvalue* or a *characteristic root* of the matrix A if:

$$A\mathbf{w} = \lambda\mathbf{w}$$

$$(A - \lambda I)\,\mathbf{w} = 0$$

- Since $\mathbf{w}$ is non-zero, this requires linear dependence in $(A - \lambda I)$, or equivalently:

$$|A - \lambda I| = 0$$

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- Consider the 2x2 case:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}$$

$$\left| \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \right| =$$

$$=$$

$$=$$

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- Consider the 2x2 case:

$$\begin{aligned}
\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} &= \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \\
\left| \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \right| &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\
&= \\
&=
\end{aligned}$$

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- Consider the 2x2 case:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}$$

$$\begin{aligned} \left| \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \right| &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= (\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} \\ &= \end{aligned}$$

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- Consider the 2x2 case:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}$$

$$\begin{aligned} \left| \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} \right| &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= (\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} \end{aligned}$$

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- Thus:

$$\begin{aligned} |A - \lambda I| &= 0 \\ &\Updownarrow \\ \lambda^2 - (a_{11} + a_{22})\,\lambda + a_{11}a_{22} - a_{12}a_{21} &= 0 \end{aligned}$$

- The last expression is the *characteristic equation* of the matrix A.
- For a $[2 \times 2]$ matrix, this is a quadratic in the eigenvalues λ for which there are at most 2 distinct solutions, real, complex or zero
- For an $[n \times n]$ matrix, this will be an n^{th} order polynomial in the eigenvalues λ, with at most n distinct solutions.

Eigenvalues

- The eigenvalues of a matrix A are those values of λ that solve:

$$|A - \lambda I| = 0$$

- For an $[n \times n]$ matrix, this will be an n^{th} order polynomial in the eigenvalues λ, with at most n distinct solutions.
  - If there are n non-zero eigenvalues, A is invertible/has linearly independent rows/columns, and is of full rank: $\text{rank}(A) = n$
  - If there are only $q \in \{1, \dots, n-1\}$ non-zero eigenvalues, A is non-invertible, has linearly dependent columns/rows and is rank deficient: $\text{rank}(A) = q$
  - If all eigenvalues are zero, the matrix A is the $[n \times n]$ zero matrix.
- We will only consider full rank matrices in this lecture.
  - Rank deficient matrices will pop up again in cointegration, so don't forget this.

Representation of powers of full rank square matrices

- If an $[n \times n]$ matrix A has n non-zero eigenvalues, there exists an invertible matrix T such that (the eigenvalue decomposition):

$$A = T\Lambda T^{-1}$$

- Where Λ is a matrix with the n non-zero eigenvalues on the diagonal:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Representation of powers of full rank square matrices

- If an $[n \times n]$ matrix A has n non-zero eigenvalues, there exists an invertible matrix T such that (the eigenvalue decomposition):

$$A = T\Lambda T^{-1}$$

- Moreover:

$$\begin{aligned} A^2 = AA &= T\Lambda T^{-1} T\Lambda T^{-1} \\ &= T\Lambda^2 T^{-1} \end{aligned}$$

- Where the standard result holds:

$$\Lambda^2 = \begin{bmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^2 \end{bmatrix}$$

Representation of powers of full rank square matrices

- If an $[n \times n]$ matrix A has n non-zero eigenvalues, there exists an invertible matrix T such that (the eigenvalue decomposition):

$$A = T\Lambda T^{-1}$$

- Moreover:

$$\begin{aligned} A^2 = AA &= T\Lambda T^{-1} T\Lambda T^{-1} \\ &= T\Lambda^2 T^{-1} \end{aligned}$$

- Where the standard result holds:

$$\Lambda^2 = \begin{bmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^2 \end{bmatrix}$$

- Thus:

$$A^i = T\Lambda^i T^{-1}$$

Representation of powers of full rank square matrices

$$A^i = T\Lambda^i T^{-1}$$

- If an $[n \times n]$ matrix A has n non-zero eigenvalues, each less than one in absolute value:

$$\lim_{i\to\infty} A^i = 0$$

Stationarity of a VAR(1)

- Returning to the backward iteration:

$$
\begin{aligned}
\mathbf{x}_t &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \\
&= A_0 + A_1 \left(A_0 + A_1 \mathbf{x}_{t-2} + \boldsymbol{e}_{t-1} \right) + \boldsymbol{e}_t \\
&= A_0 + A_0 A_1 + \boldsymbol{e}_t + A_1 \boldsymbol{e}_{t-1} + A_1^2 \left(A_0 + A_1 \mathbf{x}_{t-3} + \boldsymbol{e}_{t-2} \right) \\
&= A_0 + A_0 A_1 + A_0 A_1^2 + \boldsymbol{e}_t + A_1 \boldsymbol{e}_{t-1} + A_1^2 \boldsymbol{e}_{t-2} + A_1^3 \left(A_0 + A_1 \mathbf{x}_{t-4} + \boldsymbol{e}_{t-3} \right) \\
&= A_0 \sum_{i=0}^{\infty} A_1^i + \sum_{i=0}^{\infty} A_1^i \boldsymbol{e}_{t-i} + \lim_{i \to \infty} A_1^i \mathbf{x}_{t-i}
\end{aligned}
$$

- By the results above, iff the eigenvalues of matrix A_1 are less than one in absolute value:

$$
\begin{aligned}
\mathbf{x}_t &= A_0 + A_1 \mathbf{x}_{t-1} + \boldsymbol{e}_t \\
&= A_0 \sum_{i=0}^{\infty} A_1^i + \sum_{i=0}^{\infty} A_1^i \boldsymbol{e}_{t-i} < \infty
\end{aligned}
$$

Stationarity of a VAR(1)

- Using Lag operator:

$$\begin{aligned}\mathbf{x}_t &= A_0 + A_1\mathbf{x}_{t-1} + e_t \\ &= A_0 + A_1 L\mathbf{x}_t + e_t \\ (I - A_1 L)\,\mathbf{x}_t &= A_0 + e_t\end{aligned}$$

- The object $(I - A_1 L)$ is the *inverse characteristic matrix polynomial* of the VAR(1) process, and the process is stationary if the characteristic roots of this polynomial are all *larger* than 1 in absolute value
- Then the inverse $(I - A_1 L)^{-1}$ is well defined and implies that the VAR(1) has an Infinite order Vector Moving Average $VMA(\infty)$ representation

$$\begin{aligned}\mathbf{x}_t &= A_0 + A_1\mathbf{x}_{t-1} + e_t \\ &= A_0\sum_{i=0}^{\infty} A_1^i + \sum_{i=0}^{\infty} A_1^i e_{t-i} < \infty\end{aligned}$$

Extending to a VAR(p)

- Consider the generic VAR(p) process:

$$\mathbf{x}_t = A_0 + A_1\mathbf{x}_{t-1} + \cdots + A_p\mathbf{x}_{t-p} + e_t$$

- We can extend the rule we just derived by a simple recasting of the process into a more complicated VAR(1) process
- Define the following objects:

$$\underset{[np\times 1]}{\mathbf{X}_t} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}, \underset{[np\times 1]}{\mathbf{D}_0} = \begin{bmatrix} A_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \underset{[np\times np]}{\mathbf{D}_1} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \underset{[np\times 1]}{\mathbf{U}_t} = \begin{bmatrix} e_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Extending to a VAR(p)

- Consider the generic VAR(p) process: $\mathbf{x}_t = A_0 + A_1\mathbf{x}_{t-1} + \cdots + A_p\mathbf{x}_{t-p} + \mathbf{e}_t$

- Define the following objects:

$$\underset{[np\times 1]}{\mathbf{X}_t} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}, \underset{[np\times 1]}{\mathbf{D}_0} = \begin{bmatrix} A_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \underset{[np\times np]}{\mathbf{D}_1} = \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & I & 0 \end{bmatrix}, \underset{[np\times 1]}{\mathbf{U}_t} = \begin{bmatrix} \mathbf{e}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- The VAR(p) is then equivalently cast into *companion form* as:

$$\mathbf{X}_t = \mathbf{D}_0 + \mathbf{D}_1\mathbf{X}_{t-1} + \mathbf{U}_t$$

- Thus the VAR(p) is stationary as long as the eigenvalues of matrix D_1 are less than one in absolute value.
- In sum, the rule stays the same as in the univariate case. There are characteristic polynomials whose roots determine stationarity

Plan

- Vector Autoregression
  - Structural vs Reduced form eq'ns
  - The Identification Problem
    - The closest estimable version of a fully general linear structural model is under-identified.
    - To identify structural innovations, the structural model must be restricted in some way. This is called an *identification strategy*.
    - The Choleski decomposition imposes a timing assumption by restricting contemporaneous effects
  - Stationarity
    - The rule for stationarity of a VAR extends directly from the univariate rule:
    - The eigenvalues of the inverse characteristic matrix equation must be outside the unit circle
  - **Analyzing the information in a VAR**
  - Back to Identification – Various approaches
  - Estimation Methods

comment

- There is a lot of detail that will take time, effort and practice to internalize
- Focus on each of the aspects in the VAR lecture plan and get the essential point of each first
- Combined with applied tutorial, this will be sufficient for evaluation
- The rest of the detail is to make you aware of many aspects that will become interesting/important/sensible only when you need to apply the methods

Summarizing the information in a VAR

- Impulse Response functions
- Variance Decompositions

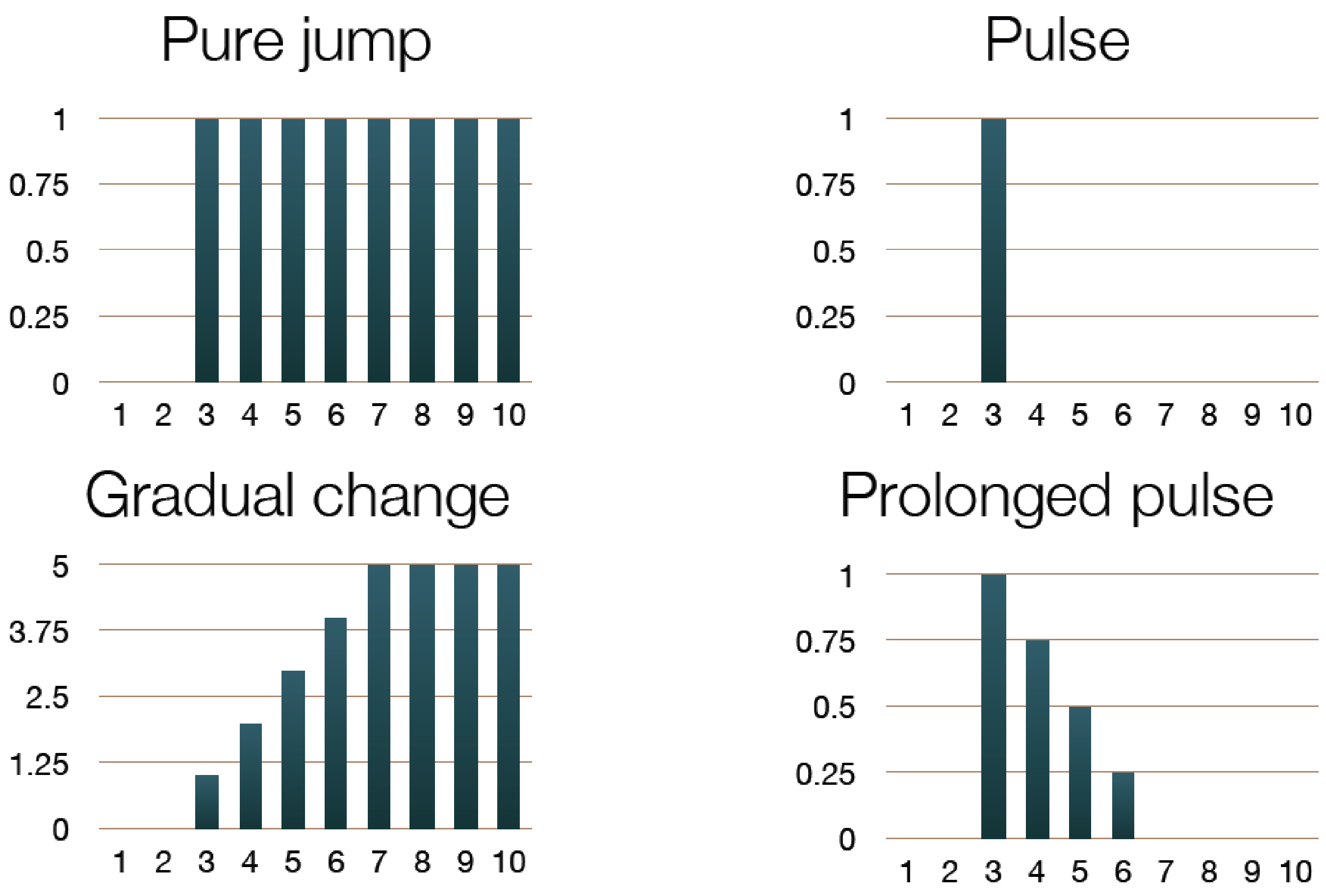
Impulse Response Function

- Starting all processes at their mean,
- What is the predicted **time path** of each variable in response to

  an "impulse"

  in each of the innovations?

Some possible Impulse Patterns



Impulse Response Functions

- In words: how does the time path of an endogenous variable respond to an innovation in one of the variables
- Theoretical IRF is w.r.t an **Innovation**
  - Structural model shock
- Empirical IRF is w.r.t a
  - Residual (no restrictions imposed), or
  - Estimate of Innovation impact based on specific identification strategy.
    - Identification strategy may influence shape of IRF, hence the economic meaning of results

Impulse Response Functions - Univariate

- The definition of an IRF is clearest in the univariate case
- Consider the simplest case: a stationary AR(1) process:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t$$

- Suppose the process starts at its expected value $y_0 = E(y_t) = \frac{a_0}{1-a_1} = \mu$, and there are no shocks (no impulse): i.e. $\varepsilon_t = 0 \; \forall \; t$

$$y_0 = \frac{a_0}{1 - a_1}$$

$$y_1 = a_0 + a_1 \left(\frac{a_0}{1 - a_1} \right)$$

$$y_1 = \frac{a_0 - a_0 a_1}{1 - a_1} + \frac{a_0 a_1}{1 - a_1}$$

$$y_1 = \frac{a_0}{1 - a_1}$$

Impulse Response Functions - Univariate

- Suppose the process starts at its expected value, $y_0 = E(y_t) = \frac{a_0}{1-a_1}$ and there is only a unit shock in period 1: $\varepsilon_1 = 1, \varepsilon_t = 0 \; \forall \, t > 1$

$$y_0 = \frac{a_0}{1 - a_1}$$

$$y_1 = a_0 + a_1 \left(\frac{a_0}{1 - a_1} \right) + \varepsilon_1 = \frac{a_0}{1 - a_1} + 1$$

$$y_2 = a_0 + a_1 \left(\frac{a_0}{1 - a_1} + 1 \right) + \varepsilon_2 = \frac{a_0}{1 - a_1} + a_1 + 0$$

$$y_3 = a_0 + a_1 \left(\frac{a_0}{1 - a_1} + a_1 \right) + \varepsilon_3 = \frac{a_0}{1 - a_1} + a_1^2 + 0$$

$$y_4 = a_0 + a_1 \left(\frac{a_0}{1 - a_1} + a_1^2 \right) + \varepsilon_3 = \frac{a_0}{1 - a_1} + a_1^3 + 0$$

Impulse Response Functions - Univariate

- Suppose the process starts at its expected value, $y_0 = E(y_t) = \frac{a_0}{1-a_1}$ and there is only a unit shock in period 1: $\varepsilon_1 = 1, \varepsilon_t = 0 \; \forall \; t > 1$

$$
\begin{aligned}
&y_0 - E(y_t) = 0 \\
&y_1 - E(y_t) = 1 \\
&y_2 - E(y_t) = a_1 \\
&y_3 - E(y_t) = a_1^2 \\
&y_4 - E(y_t) = a_1^3 \\
&\ldots \\
&y_t - E(y_t) = a_1^{t-1}
\end{aligned}
$$

- This sequence is the Impulse Response Function for this process given a once-off pulse in the error sequence

Impulse Response Functions - Univariate

- This generalizes to any order stationary ARMA(p,q) process:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \cdots + b_q \varepsilon_{t-q}$$

- We can always represent a stationary ARMA(p,q) process as an MA(∞) process:

$$y_t = \mu + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

  – With $c_0 = 1$

- Thus:

$$y_t - \mu = \varepsilon_t + c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \cdots$$

Impulse Response Functions - Univariate

- We can always represent a stationary ARMA(p,q) process as an MA(\infty) process:

$$y_t = \mu + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i}$$

- Thus:

$$y_t - \mu = \varepsilon_t + c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \cdots$$

- If $\{\varepsilon_t\}_{t=1}^{T} = [1,0,0,\dots]$, then:

$$\begin{aligned} y_1 - \mu &= \varepsilon_1 + c_1 \varepsilon_0 + c_2 \varepsilon_{-1} + \cdots \\ y_1 - \mu &= 1 + c_1 0 + c_2 0 + \cdots = 1 \\ y_2 - \mu &= 0 + c_1 1 + c_2 0 + \cdots = c_1 \\ y_3 - \mu &= 0 + c_1 0 + c_2 1 + \cdots = c_2 \end{aligned}$$

- The IRF for this process is:

$$[1, c_1, c_2, c_3, \dots]$$

- This extends directly to the multivariate case, except that if we have a jointly endogenous process with n shocks, there are n IRFs for each variable

IRF - derivation

- Rewrite reduced form with lag operators:

$$
\begin{aligned}
\mathbf{x}_t &= \mathbf{A}_0 + \mathbf{A}_1 \mathbf{x}_{t-1} + \ldots + \mathbf{A}_p \mathbf{x}_{t-p} + e_t \\
&= \mathbf{A}_0 + \mathbf{A}_1 L \mathbf{x}_t + \ldots + \mathbf{A}_p L^p \mathbf{x}_t + e_t \\
&= \mathbf{A}_0 + \mathbf{A}\left(L\right) \mathbf{x}_t + e_t \\
\left(\mathbf{I} - \mathbf{A}\left(L\right)\right) \mathbf{x}_t &= \mathbf{A}_0 + e_t
\end{aligned}
$$

- $\left(\mathbf{I} - \mathbf{A}\left(L\right)\right)$ is the inverse characteristic matrix polynomial of this process
- If the process is stationary, its inverse exists

IRF - derivation

- If the process is stationary:

$$
\begin{aligned}
(\mathbf{I}-\mathbf{A}(L))\,\mathbf{x}_t &= \mathbf{A}_0 + e_t \\
\mathbf{x}_t &= (\mathbf{I}-\mathbf{A}(L))^{-1}(\mathbf{A}_0 + e_t) \\
&= (\mathbf{I}-\mathbf{A}(L))^{-1}\mathbf{A}_0 + (\mathbf{I}-\mathbf{A}(L))^{-1} e_t \\
&= (\mathbf{I}-\mathbf{A}(L))^{-1}\mathbf{A}_0 + (\mathbf{I}-\mathbf{A}(L))^{-1}\mathbf{B}^{-1}\varepsilon_t \\
&= \boldsymbol{\mu} + \mathbf{C}(L)\,\varepsilon_t
\end{aligned}
$$

- Where:

$$\boldsymbol{\mu} = (\mathbf{I}-\mathbf{A}(1))^{-1}\mathbf{A}_0 = E[\mathbf{x}_t]$$

IRF - derivation

- Unpacking the representation result:

$$\begin{aligned}
\mathbf{x}_t - \boldsymbol{\mu} &= \mathbf{C}(L)\,\varepsilon_t \\
&= \mathbf{C}^{(0)}\varepsilon_t + \mathbf{C}^{(1)}\varepsilon_{t-1} + \mathbf{C}^{(2)}\varepsilon_{t-2} + \ldots \\
&= \sum_{i=0}^{\infty} \mathbf{C}^{(i)}\varepsilon_{t-i} \\
&= \left(\sum_{i=0}^{\infty} \mathbf{C}^{(i)} L^i\right)\varepsilon_t
\end{aligned}$$

IRF - derivation

- Back to two variable case:

$$\mathbf{x}_t - \boldsymbol{\mu} = \left[\begin{array}{c} y_t - \mu_y \\ z_t - \mu_z \end{array}\right] = \left[\begin{array}{cc} c_{11}^{(0)} & c_{12}^{(0)} \\ c_{21}^{(0)} & c_{22}^{(0)} \end{array}\right] \left[\begin{array}{c} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{array}\right] + \left[\begin{array}{cc} c_{11}^{(1)} & c_{12}^{(1)} \\ c_{21}^{(1)} & c_{22}^{(1)} \end{array}\right] \left[\begin{array}{c} \varepsilon_{y,t-1} \\ \varepsilon_{z,t-1} \end{array}\right] + \ldots$$

- The IRF is the impact on the time paths of the endogenous variables of a sequence of shocks:
  - Let the impulse be: $\varepsilon_{y,1} = 1$ and $\varepsilon_{z,1} = \varepsilon_{y,t} = 0 \; \forall \; t > 1$
  - Both variables respond:

$$y_1 - \mu_y = \varepsilon_{y,1} + c_{11}^{(0)} \varepsilon_{y,0} + c_{11}^{(1)} \varepsilon_{y,-1} + \cdots$$
$$z_1 - \mu_z = \varepsilon_{y,1} + c_{12}^{(0)} \varepsilon_{y,0} + c_{12}^{(1)} \varepsilon_{y,-1} + \cdots$$

IRF - derivation

- Back to two variable case:

$$\mathbf{x}_t - \boldsymbol{\mu} = \left[\begin{array}{c} y_t - \mu_y \\ z_t - \mu_z \end{array}\right] = \left[\begin{array}{cc} c_{11}^{(0)} & c_{12}^{(0)} \\ c_{21}^{(0)} & c_{22}^{(0)} \end{array}\right]\left[\begin{array}{c} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{array}\right] + \left[\begin{array}{cc} c_{11}^{(1)} & c_{12}^{(1)} \\ c_{21}^{(1)} & c_{22}^{(1)} \end{array}\right]\left[\begin{array}{c} \varepsilon_{y,t-1} \\ \varepsilon_{z,t-1} \end{array}\right] + \dots$$

- The IRF is the impact on the time paths of the endogenous variables of a sequence of shocks:
  - Let the impulse be: $\varepsilon_{y,1} = 1$ and $\varepsilon_{z,1} = \varepsilon_{y,t} = 0 \;\forall\; t > 1$

$$\left\{\boldsymbol{\varepsilon}_{t+s}^{temporary}\right\}_{s=0}^{T} = \left[\begin{array}{cccc} 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \end{array}\right]$$

$$E_t\left[\left\{\mathbf{x}_{t+s} - \boldsymbol{\mu}\right\}_{s=0}^{T}\middle|\left\{\boldsymbol{\varepsilon}_t^{temporary}\right\}\right] = \left[\begin{array}{cccc} c_{11}^{(0)} & c_{11}^{(1)} & c_{11}^{(2)} & \dots \\ c_{21}^{(0)} & c_{21}^{(1)} & c_{21}^{(2)} & \dots \end{array}\right]$$

IRF - derivation

- $\mathbf{C}(L)$ is a convergent matrix valued polynomial in the lag operator
- What happens if we substitute L with 1?

$$\mathbf{C}(L) = \mathbf{C}^{(0)} + \mathbf{C}^{(1)}L + \mathbf{C}^{(2)}L^2 + \cdots$$
$$\mathbf{C}(1) = \mathbf{C}^{(0)} + \mathbf{C}^{(1)}1 + \mathbf{C}^{(2)}1^2 + \cdots$$

- Convergent means the [n x n] matrix:

$$\mathbf{C}(1) = \sum_{i=0}^{\infty} \mathbf{C}^{(i)}$$

has only finite valued entries

Two distinct processes:

- Process 1
  - No contemporaneous effect of x_{2t} on x_{1t}:

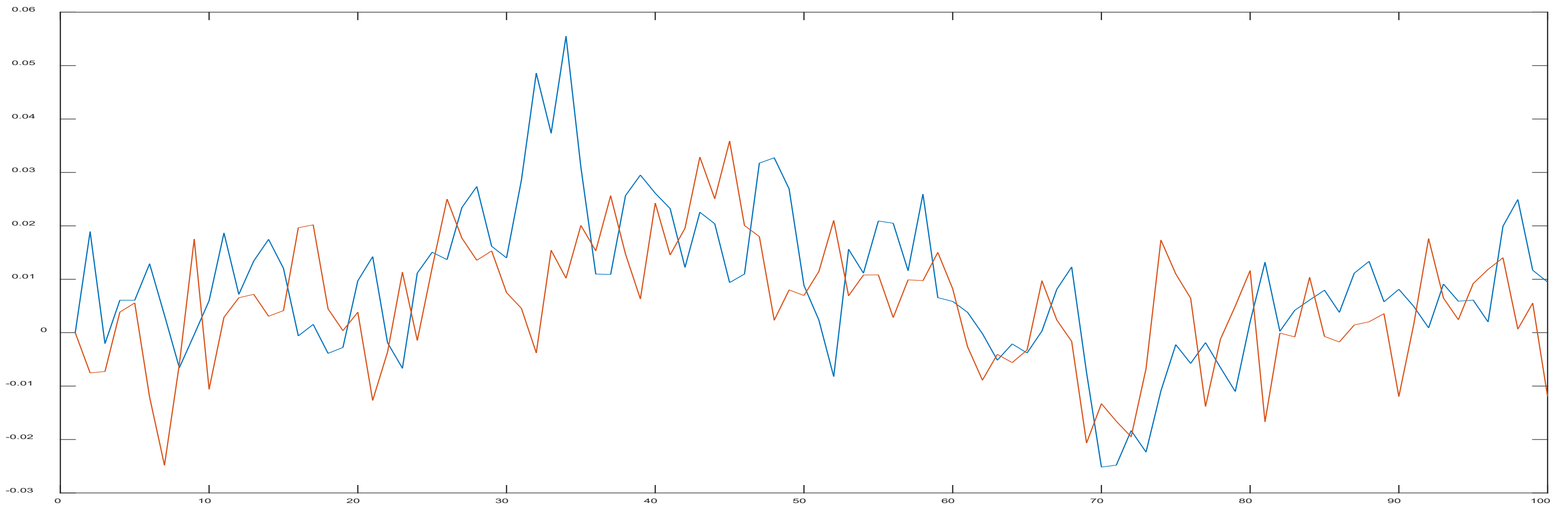
$$\begin{bmatrix} 1 & 0 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

- Process 2
  - Mutual contemporaneous effects:

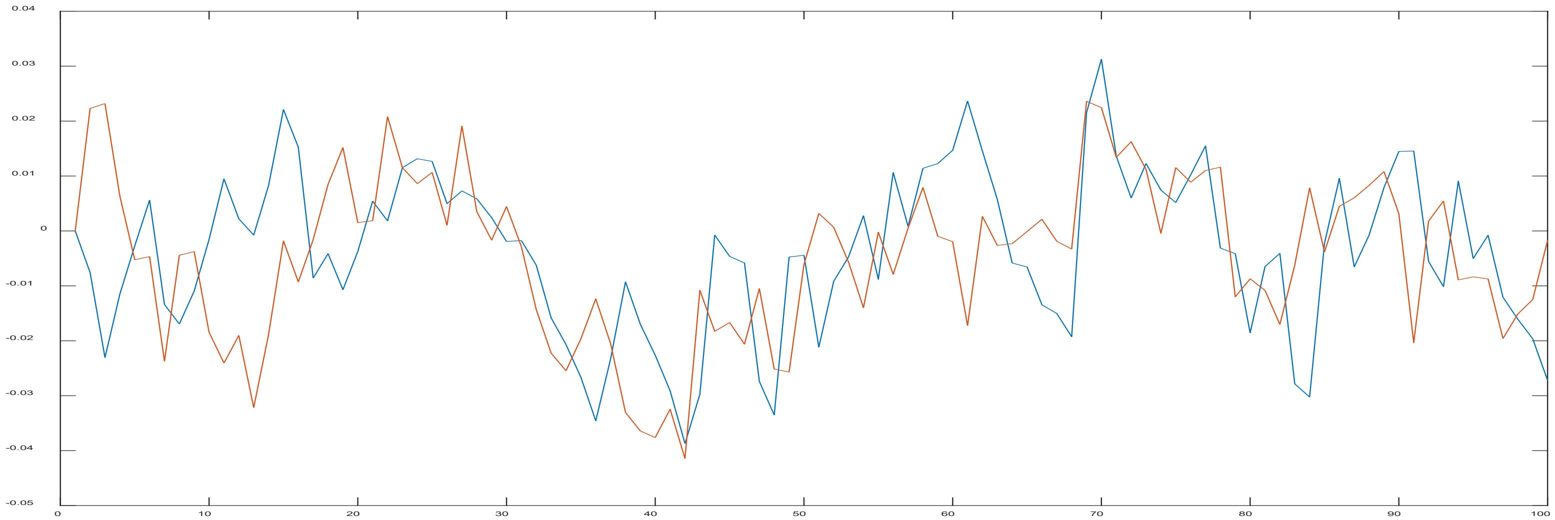
$$\begin{bmatrix} 1 & 0.1 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Two independent processes:

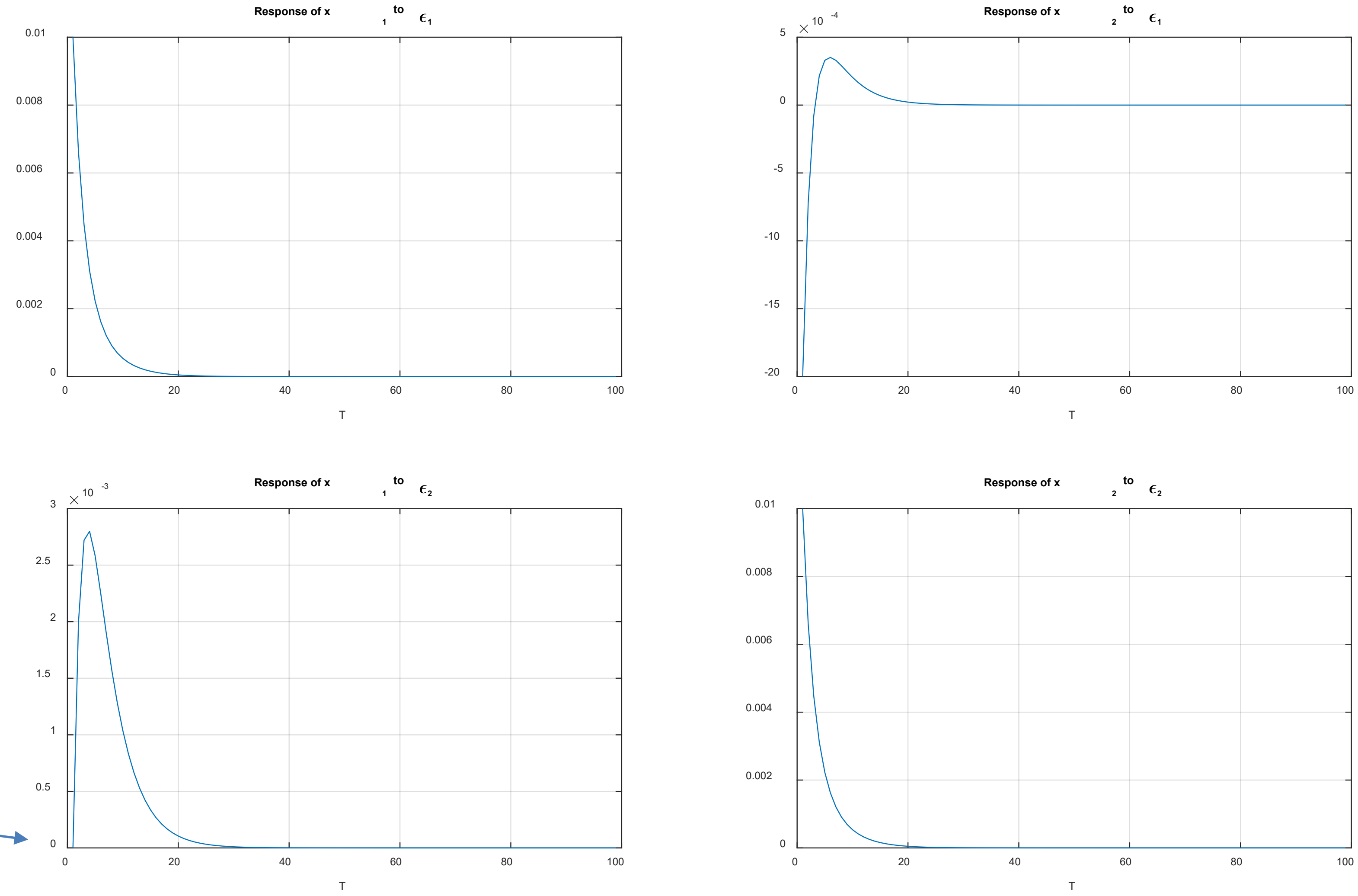
Process 1



Process 2

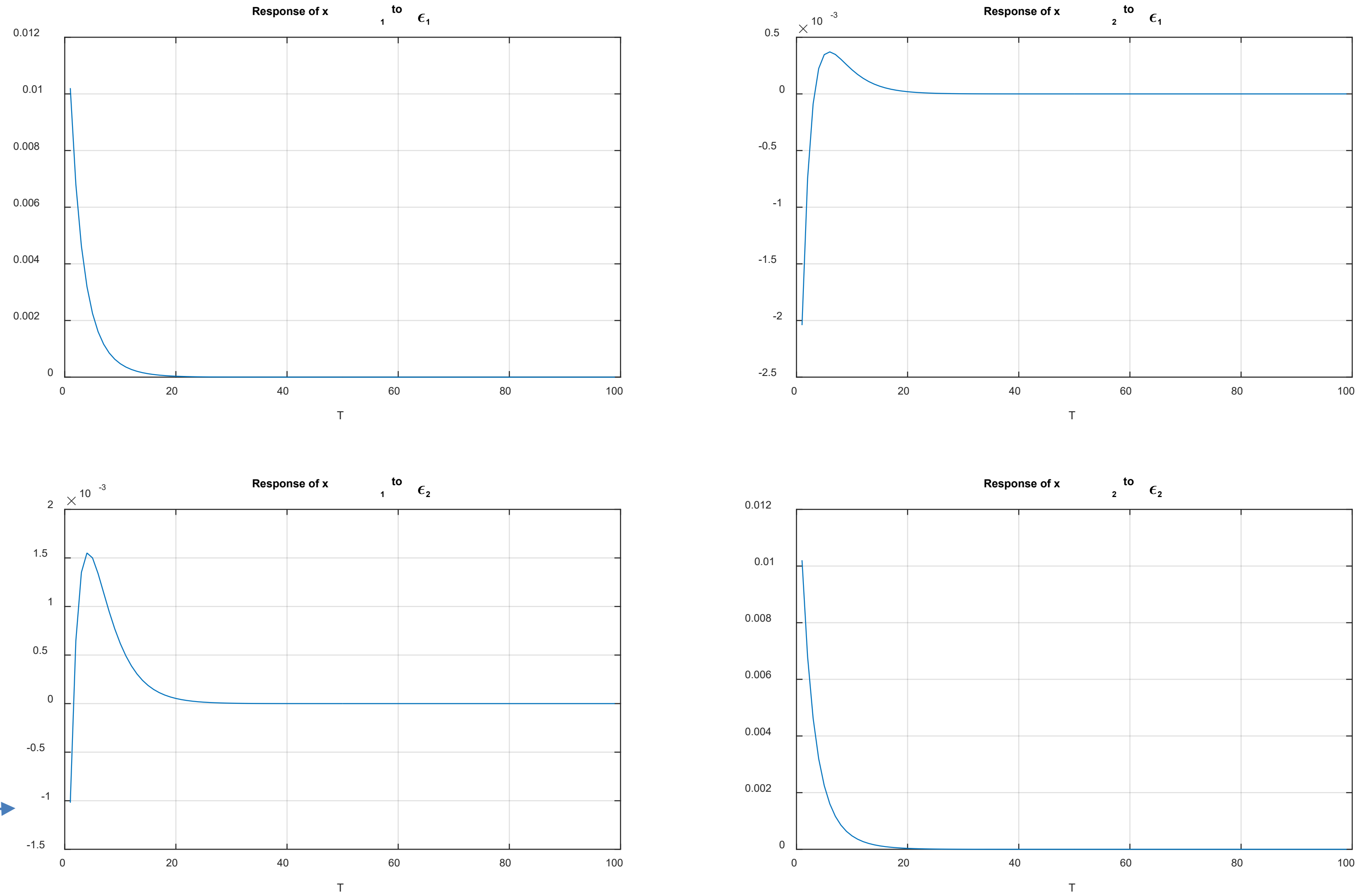


Theoretical IRF of Process 1: $\begin{bmatrix} 1 & 0 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$



Note: no immediate response of x_{1t} to ε_{2t}

Theoretical IRF of Process 2: $\begin{bmatrix} 1 & 0.1 \\ 0.2 & 1 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} = \begin{bmatrix} 0.7 & 0.2 \\ 0.2 & 0.7 \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$



Note: immediate response of x_{1t} to ε_{2t}

Two versions of impulse responses

- Since we are considering stationary VARs, all processes must eventually go back to their respective expected values
  - This means that IRFs from a once off shock must return to zero
  - This is provides a visual "test" – if the impulse does not return to zero fast enough, suspect non-stationarity
- We might also be interested in the cumulative effect of a shock
  - Then we can construct the **cumulative IRF**
  - Equivalent to an impulse response to a permanent step change
  - If the IRF is $[1, c_1, c_2, c_3, \ldots]$,
    the cumulative IRF is $[1, 1 + c_1, 1 + c_1 + c_2, 1 + c_1 + c_2 + c_3, \ldots]$

Summarizing the information in a VAR

- Impulse Response functions
- **Variance Decompositions**
  - Unlike the impulse response function, the variance decomposition is not interesting in the univariate case
  - A more complete name is *Forecast Error Variance Decomposition*
  - In the multivariate case this is interesting:
    - Given that a set of variables are jointly endogenous, the uncertainty in forecasts in an individual variable is due to *all* the shocks in the system of equations
    - The variance decomposition determines how much is due to which shock for every forecast horizon

Variance Decomposition

- Consider the process in period t + n in VMA form:

$$\mathbf{x}_{t+n} - \boldsymbol{\mu} = \sum_{i=0}^{\infty} \mathbf{C}^{(i)} \boldsymbol{\varepsilon}_{t+n-i}$$

- The Expected value conditional on period t information is:

$$E_t\left[\mathbf{x}_{t+n}\right] = \boldsymbol{\mu} + \sum_{i=n}^{\infty} \mathbf{C}^{(i)} \boldsymbol{\varepsilon}_{t+n-i}$$

- Note that the only difference is in the indices of the summation
  - The expected value of future shocks is zero

Variance Decomposition

- This defines the n-period ahead **Forecast Error**:

$$\mathbf{x}_{t+n} - E_t\left(\mathbf{x}_{t+n}\right) = \sum_{i=0}^{n-1} \mathbf{C}^{(i)} \boldsymbol{\varepsilon}_{t+n-i}$$

- Extracting the row that corresponds to y:

$$\begin{aligned} y_{t+n} - E_t\left(y_{t+n}\right) \quad &= \quad c_{11}^{(0)} \varepsilon_{y,t+n} + c_{11}^{(1)} \varepsilon_{y,t+n-1} + \ldots + c_{11}^{(n-1)} \varepsilon_{y,t+1} \\ & \qquad + c_{12}^{(0)} \varepsilon_{z,t+n} + c_{12}^{(1)} \varepsilon_{z,t+n-1} + \ldots + c_{12}^{(n-1)} \varepsilon_{z,t+1} \end{aligned}$$

- Note: uncertainty grows as we go further into the future

Variance Decomposition

- The **variance decomposition** is the proportions of this uncertainty that is due to each structural shock
- The n-period ahead total forecast error variance of the y process is:

$$
\begin{aligned}
\varsigma_y^2(n) &= E_t\left[y_{t+n} - E_t\left(y_{t+n}\right)\right]^2 \\
&= \sigma_y^2\left[\left(c_{11}^{(0)}\right)^2 + \left(c_{11}^{(1)}\right)^2 + \ldots + \left(c_{11}^{(n-1)}\right)^2\right] \\
&\quad + \sigma_z^2\left[\left(c_{12}^{(0)}\right)^2 + \left(c_{12}^{(1)}\right)^2 + \ldots + \left(c_{12}^{(n-1)}\right)^2\right]
\end{aligned}
$$

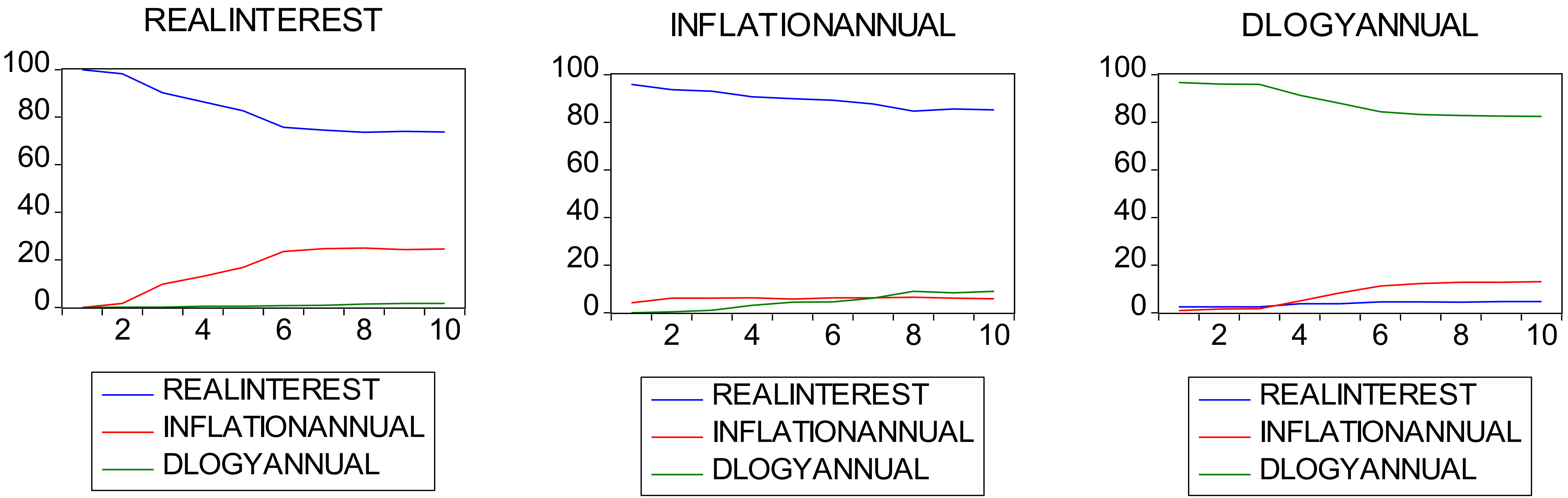
Variance Decomposition

- Thus we can split $\varsigma_y^2(n)$ into two parts:

$$\text{due to } \varepsilon_{y,t}\text{:} \quad \frac{\sigma_y^2\left[\left(c_{11}^{(0)}\right)^2 + \left(c_{11}^{(1)}\right)^2 + \ldots + \left(c_{11}^{(n-1)}\right)^2\right]}{\varsigma_y^2(n)}$$

$$\text{due to } \varepsilon_{z,t}\text{:} \quad \frac{\sigma_z^2\left[\left(c_{12}^{(0)}\right)^2 + \left(c_{12}^{(1)}\right)^2 + \ldots + \left(c_{12}^{(n-1)}\right)^2\right]}{\varsigma_y^2(n)}$$

Example:


REALINTEREST
INFLATIONANNUAL
DLOGYANNUAL
100
80
60
40
20
0
2
4
6
8
10
REALINTEREST
INFLATIONANNUAL
DLOGYANNUAL

Plan

- Vector Autoregression
  - Structural vs Reduced form eq’ns
  - The Identification Problem
  - Stationarity
  - Analyzing the information in a VAR
  - **Back to Identification – Various approaches**
  - Estimation Methods

Identification

- We need $\frac{n^2-n}{2}$ restrictions to identify the structural innovations/IRFs

Structural VAR

- As soon as we use economic theory to restrict parameters in a VAR, it becomes a “structural VAR”

- Economic theory might suggest more than the necessary identification restrictions
  - Additional restrictions will worsen the in-sample fit
  - The degree to which this happens can be used as a test of the additional restrictions
  - These are commonly called *over-identifying* restrictions

Choleski Decomposition

- In the unrestricted primitive form,

$$B\mathbf{x}_t = \Gamma_0 + \Gamma_1\mathbf{x}_{t-1} + \cdots + +\Gamma_p\mathbf{x}_{t-p} + \varepsilon_t$$

- The Choleski decomposition implies choosing a temporal ordering:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -b_{21} & 1 & 0 & \cdots & 0 \\ -b_{31} & -b_{32} & 1 & \vdots & \vdots \\ \vdots & \vdots & \cdots & \ddots & 0 \\ -b_{n1} & -b_{n2} & \cdots & -b_{n,n-1} & 1 \end{bmatrix}$$

Variance Restrictions

- We have shown:

$$\begin{aligned}\boldsymbol{e}_t &= C^{(0)}\boldsymbol{\varepsilon}_t \\ &= \begin{bmatrix} c_{11}^{(0)} & c_{12}^{(0)} \\ c_{21}^{(0)} & c_{22}^{(0)} \end{bmatrix} \begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{z,t} \end{bmatrix}\end{aligned}$$

- From this we can obtain three (non-linear) equations in 4 unknowns

Variance Restrictions

- From this we can obtain three (non-linear) equations in 4 unknowns

$$
\begin{aligned}
E\left(\boldsymbol{e}_t \boldsymbol{e}_t'\right) &= E\left(C^{(0)} \boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' C^{(0)\prime}\right) \\
\begin{bmatrix} \sigma_{e_1}^2 & \sigma_{e_1 e_2} \\ \sigma_{e_1 e_2} & \sigma_{e_2}^2 \end{bmatrix} &= C^{(0)} I C^{(0)\prime} \\
&= C^{(0)} C^{(0)\prime}
\end{aligned}
$$

In this identification scheme, we can know nothing about the primitive variances, so we can just set them to 1

Variance Restrictions

- From this we can obtain three (non-linear) equations in 4 unknowns

$$
\begin{aligned}
\sigma_{e_1}^2 &= \left(c_{11}^{(0)}\right)^2 + \left(c_{12}^{(0)}\right)^2 \\
\sigma_{e_2}^2 &= \left(c_{21}^{(0)}\right)^2 + \left(c_{22}^{(0)}\right)^2 \\
\sigma_{e_1 e_2} &= c_{11}^{(0)} c_{21}^{(0)} + c_{12}^{(0)} c_{22}^{(0)}
\end{aligned}
$$

Variance Restrictions

- From this we can obtain three (non-linear) equations in 4 unknowns
- An additional restriction will allow us (potentially) to recover the C(0) coefficients and hence the structural shocks
- Rarely used in isolation in practice = no strong ex ante reason to restrict variances.

Other Identification Approaches:

- Blanchard-Quah decomposition
- Uhlig's Sign Restrictions on IRF
- Pesaran and Shin's Generalized IRF

- First an understanding check: What will the impact of different choices for a *just-identified* system be on forecasting?

Blanchard – Quah Identification Strategy

- Extends the Nelson and Plosser (1982) exercise that split the variation in GDP in permanent and temporary components
  - Showed that most of the variation can be explained by permanent shocks
  - Beginning of the Real Business Cycle literature
  - No unique way of doing this
- Blanchard and Quah (1989) use variance relations combined with a long run impact restrictions to identify a permanent and a temporary innovation

Blanchard – Quah Identification Strategy

- Blanchard and Quah (1989) long run impact restrictions:
- Recall that a stationary VAR is equivalent to a convergent series of coefficients in the VMA representation:

$$\mathbf{C}(1) = \sum_{i=0}^{\infty} \mathbf{C}^{(i)}$$

- The elements of this matrix may still be positive (or negative) so that a sequence of innovations may have a permanent cumulative effect
- The BQ approach involves setting one or more of the entries in $\mathrm{C}(1)$ to **zero**
  - Which implies the assumption that of these shocks have no permanent effect on the relevant variables

Blanchard – Quah Identification Strategy

- Blanchard and Quah (1989) setup:
- Endogenous Variables:

  Log real GDP - y_t - unit root, non stationary

  Growth in real GDP - Δy_t stationary

  Unemployment - u_t stationary

- Structural Innovations:

  Demand shock - $\varepsilon_{d,t}$ - temporary effects on ALL variables

  Supply shock - $\varepsilon_{s,t}$ - permanent effect on y_t

Blanchard – Quah Identification Strategy

- Blanchard and Quah (1989) setup:
- VAR in VMA form:

$$\mathbf{x}_t - \mu = \begin{bmatrix} \Delta y_t - \mu_{\Delta y} \\ u_t - \mu_u \end{bmatrix} = \begin{bmatrix} c^{(0)}_{\Delta y,d} & c^{(0)}_{\Delta y,s} \\ c^{(0)}_{u,d} & c^{(0)}_{u,s} \end{bmatrix} \begin{bmatrix} \varepsilon_{d,t} \\ \varepsilon_{s,t} \end{bmatrix} + \begin{bmatrix} c^{(1)}_{\Delta y,d} & c^{(1)}_{\Delta y,s} \\ c^{(1)}_{u,d} & c^{(1)}_{u,s} \end{bmatrix} \begin{bmatrix} \varepsilon_{d,t-1} \\ \varepsilon_{s,t-1} \end{bmatrix} + \ldots$$

- Demand shock has no long run impact for a given sequence of shocks on y if:

$$\sum_{k=0}^{\infty} c^{(k)}_{\Delta y,d} \varepsilon_{d,t-k} = 0$$

- For all possible sequences of shocks:

$$\sum_{k=0}^{\infty} c^{(k)}_{\Delta y,d} = 0$$

Pesaran and Shin’s generalized IRF

Pesaran and Shin's generalized IRF

- Recall our original primitive model:

$$\underset{[n\times n]}{\mathbf{B}}\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\mathbf{\Gamma}_0} + \underset{[n\times n]}{\mathbf{\Gamma}_1}\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \mathbf{\Gamma}_p\mathbf{x}_{t-p} + \varepsilon_t \qquad E\left[\varepsilon_t\varepsilon_t'\right] = \underset{[n\times n]}{\mathbf{\Sigma}_\varepsilon}$$

  - Where:
    - B is unrestricted (non-diagonal, non-symmetric)
    - Σ_ε is diagonal (pure uncorrelated innovations)
- This is not directly estimable, and lead to the reduced form:

$$\underset{[n\times 1]}{\mathbf{x}_t} = \underset{[n\times 1]}{\mathbf{A}_0} + \underset{[n\times n]}{\mathbf{A}_1}\underset{[n\times 1]}{\mathbf{x}_{t-1}} + \ldots + \mathbf{A}_p\mathbf{x}_{t-p} + \boldsymbol{e}_t \qquad E\left[\mathbf{e}_t\mathbf{e}_t'\right] = \underset{[n\times n]}{\mathbf{\Sigma}}$$

- Where Σ is non-diagonal but symmetric, and "identification" was about recovering the B matrix

Pesaran and Shin's generalized IRF

- Pesaran and Shin start with the primitive model:

$$\mathbf{x}_t = \Gamma_0 + \Gamma_1 \mathbf{x}_{t-1} + \ldots + \Gamma_p \mathbf{x}_{t-p} + \varepsilon_t$$

$$\begin{aligned} E\left[\varepsilon_t \varepsilon_t'\right] = \Sigma &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \cdots & \sigma_{2n} \\ \sigma_{12} & \sigma_{23} & \sigma_{33} & \cdots & \sigma_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \sigma_{3n} & \cdots & \sigma_{nn} \end{bmatrix} \\ E\left[\varepsilon_t \varepsilon_{t+s}'\right] &= 0 \forall s \neq 0 \end{aligned}$$

  - Implying:
    - B is restricted – assumed to be identity matrix
    - Σ_ε is positive definite, symmetric
      - Fundamental errors may be correlated

Pesaran and Shin's generalized IRF

- This means that there can be no "pure impulse" to one innovation. Since they are correlated, a shock to one is a shock to all:

$$E\left(\varepsilon_t | \varepsilon_{jt} = \delta_j\right) = \left(\sigma_{1j}, \sigma_{2j}, ..., \sigma_{mj}\right)' \sigma_{jj}^{-1} \delta_j = \mathbf{\Sigma} \mathbf{e}_j \sigma_{jj}^{-1} \delta_j.$$

- Giving the GIRF as:

$$\left(\frac{\mathbf{A}_n \mathbf{\Sigma} \mathbf{e}_j}{\sqrt{\sigma_{jj}}}\right)\left(\frac{\delta_j}{\sqrt{\sigma_{jj}}}\right), \; n = 0, 1, 2, ... \; .$$

- Or normalized as:

$$\psi_j^g(n) = \sigma_{jj}^{-\frac{1}{2}} \mathbf{A}_n \mathbf{\Sigma} \mathbf{e}_j, \; n = 0, 1, 2, ...,$$

Pesaran and Shin’s generalized IRF

- Evaluation:
  - Using Pesaran and Shin’s definition allows mutual “contemporaneous” effects in a VAR
  - However, these effects are all via the estimated Var-Covar of the residuals
  - Thus necessarily symmetric
  - Cannot identify fundamental contemporaneous feedback that might be asymmetric
  - This is identification of a sort, but different from the program we began with.

Uhlig’s Sign Restricted IRF

- Motivation:
  - Sims’ unrestricted VAR approach lead to empirical “puzzles”
    - Price Puzzle: prices tend to rise after contractionary monetary policy
    - This is a “puzzle” because it does not fit theory
  - even with Cholesky ordering, some “implicit” theorizing is always done
  - Residuals from VAR estimate are always a linear combination of structural innovations

- Do the theorizing explicitly to impose “what we really think we know”
  - Force (some) IRFs to go in the “right” direction
  - Test the model by leaving IRFs of unknown response (theoretically) unrestricted.

Uhlig's Sign Restricted IRF

- Operationalized Method:
  - Brute force simulation
  - Draw (random) impulse vectors for restricted responses, impose sign restrictions
  - Fit unrestricted impulse response functions
  - Repeat 10000 times
  - Plot distribution of impulse response functions
- Details beyond this course
  - Uhlig shows that the asymptotics are reliable

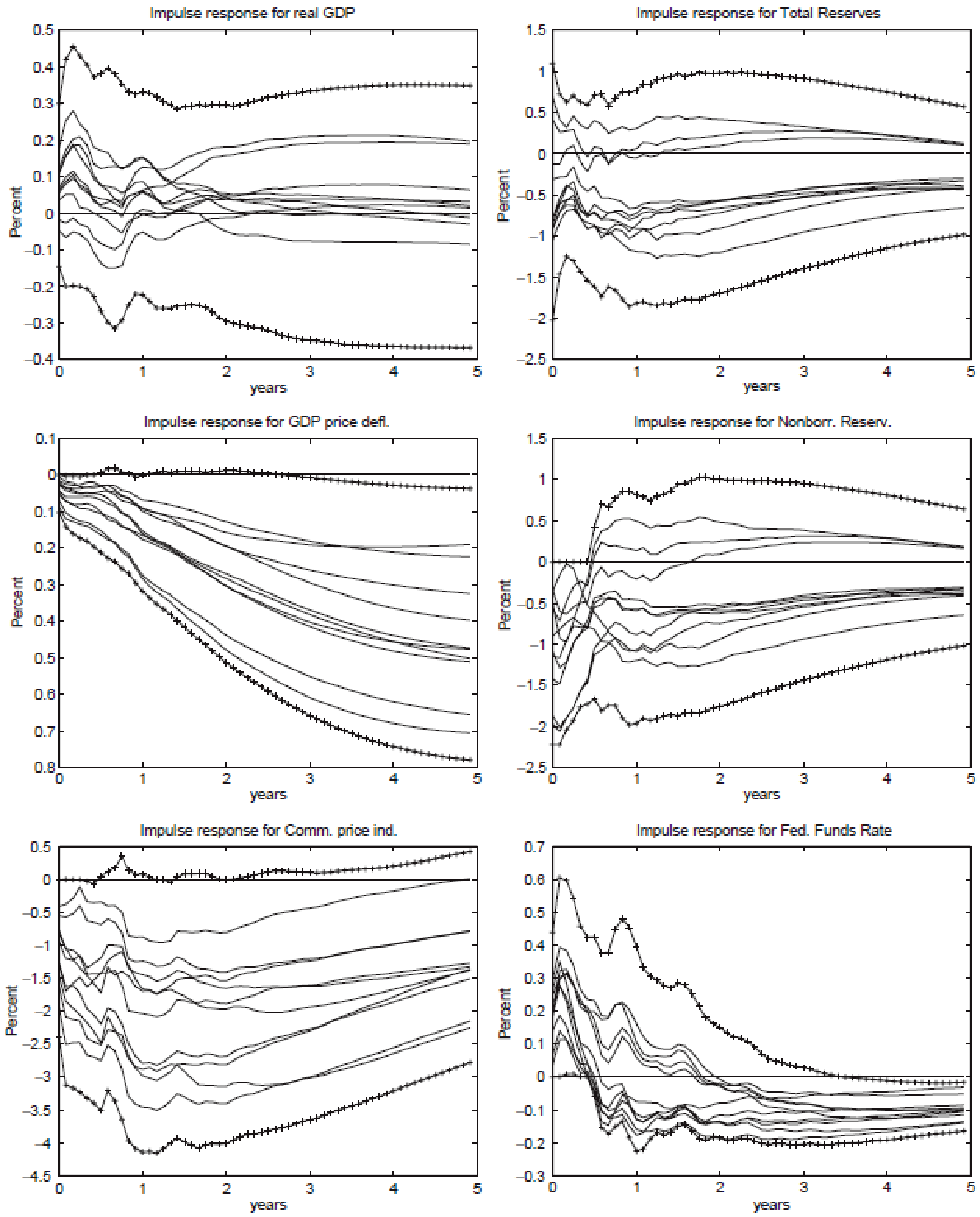


Fig. 2. This figure shows the possible range of impulse response functions when imposing the sign restrictions for $K = 5$ at the OLSE point estimate for the VAR.

Uhlig's Sign Restricted IRF

- Concerns:
  - One has to choose "how long" an IRF cannot violate the sign imposed
  - Recent studies (Wolf 2016) shows that, theoretically, it may still be impossible to know that what was identified was a monetary policy shock
  - Could be a linear combination of other shocks

Plan

- Vector Autoregression
  - Structural vs Reduced form eq’ns
  - The Identification Problem
  - Stationarity
  - Analyzing the information in a VAR
  - Back to Identification – Various approaches
  - **Estimation Methods**
  - Evaluation of Fit

Estimating a VAR

- A VAR is readily estimable via OLS – linear in parameters
- Consider the generic VAR(p) process:

$$\mathbf{x}_t = A_1\mathbf{x}_{t-1} + \cdots + A_p\mathbf{x}_{t-p} + \boldsymbol{e}_t$$

- Suppose we have T+p observations on each of the n variables
- Define:

$$\underset{[np\times 1]}{\mathbf{X}_t} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}$$

$$\underset{[n\times T]}{Y} = \begin{bmatrix} \mathbf{x}_1, & \ldots, & \mathbf{x}_T \end{bmatrix}$$

$$\underset{[np\times T]}{Z} = \begin{bmatrix} \mathbf{X}_0, & \ldots, & \mathbf{X}_{T-1} \end{bmatrix}$$

$$\underset{[n\times np]}{A} = \begin{bmatrix} A_1, & \ldots, & A_p \end{bmatrix}$$

$$\underset{[n\times T]}{U} = \begin{bmatrix} \boldsymbol{e}_1, & \ldots, & \boldsymbol{e}_T \end{bmatrix}$$

Estimating a VAR

$$\mathbf{x}_t = A_1\mathbf{x}_{t-1} + \cdots + A_p\mathbf{x}_{t-p} + \boldsymbol{e}_t$$

$$\underset{[np\times 1]}{\mathbf{X}_t} = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{x}_{t-1} \\ \vdots \\ \mathbf{x}_{t-p+1} \end{bmatrix}$$

$$\underset{[n\times T]}{Y} = \left[\begin{array}{ccc} \mathbf{x}_1, & \ldots, & \mathbf{x}_T \end{array}\right]$$

$$\underset{[np\times T]}{Z} = \left[\begin{array}{ccc} \mathbf{X}_0, & \ldots, & \mathbf{X}_{T-1} \end{array}\right]$$

$$\underset{[n\times np]}{A} = \left[\begin{array}{ccc} A_1, & \ldots, & A_p \end{array}\right]$$

$$\underset{[n\times T]}{U} = \left[\begin{array}{ccc} \boldsymbol{e}_1, & \ldots, & \boldsymbol{e}_T \end{array}\right]$$

$$Y = AZ + U$$

$$\widehat{A} = YZ'\left(ZZ'\right)^{-1}$$

Estimating a VAR

- A VAR is linear in parameters and variables, so can easily be estimated by OLS,

- Hence the estimator has the standard properties under suitable conditions:
  - Consistent (why not unbiased?)
  - Asymptotically normal

Plan

- Vector Autoregression
  - Structural vs Reduced form eq’ns
  - The Identification Problem
  - Stationarity
  - Analyzing the information in a VAR
  - Back to Identification – Various approaches
  - Estimation Methods
  - **Evaluation of Fit**

Evaluating Fit

- As in the univariate case, an adequate model must have white noise residuals
  - Check autocorrelation functions
  - Tests for serial correlation

Which Variables belong in a VAR?

- Up to now we've assumed there is some given simultaneous system we need to uncover.
  - Theory would suggest the variables

- How do we know which variables belong in a VAR?

- Standard tool:
  - Granger Causality
  - Block Exogeneity test

Which Variables belong in a VAR?

- Economic Theory is usually where we start
  - But macro models can get very large
  - Imagine simultaneously testing Purchasing Price Parity, Uncovered Interest Parity and the Term Structure hypotheses
  - This leads to the curse of dimensionality:
  - Suppose we have 5 variables and 4 lags
    - 100 slope coefficients:
    - 5 constants
    - 15 variance/covariance terms

$$\underset{[n \times np]}{A} = [\ A_1, \quad \ldots, \quad A_p\ ]$$

  - Curse of Dimensionality: require large data sets for any kind of accurate estimate
  - Typically: use a limited number of variables and lags to answer a narrow question

Which Variables belong in a VAR?

- Other approaches to deal with dimensionality problem
  - Bayesian VARs (to some extent)
  - Factor Augmented VARs

- FAVARs augment a VAR with Factors extracted from a large dataset
  - Factors can be estimated as the first principle components of a large data set, after which the VAR is estimated
  - Or jointly estimated with VAR coefficients using Bayesian methods
  - Bernanke, Boivin and Eliasz (2004) augment a basic monetary VAR with the factors extracted from 120 disaggregated variables

Which Variables belong in a VAR?

- Given a set of variables that represent some economic process of interest, we are also interested if the variables *empirically* belong in the model of the DGP

- Standard tool:
  - Granger Causality
  - Block Exogeneity test

Granger Causality Test

- This is a simple F test:

$$\mathbf{x}_t = A_0 + A(L)\,\mathbf{x}_{t-1} + \boldsymbol{e}_t$$

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} a_{01} \\ a_{02} \end{bmatrix} + \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \ldots + \begin{bmatrix} a_{11}^{(p)} & a_{12}^{(p)} \\ a_{21}^{(p)} & a_{22}^{(p)} \end{bmatrix} \begin{bmatrix} y_{t-p} \\ z_{t-p} \end{bmatrix} + \begin{bmatrix} e_{y,t} \\ e_{z,t} \end{bmatrix}$$

The null-hypothesis that z_t does not Granger-cause y_t is:

$$H_0 : \; a_{12}^{(1)} = a_{12}^{(2)} = \ldots = a_{12}^{(p)} = 0$$

Correct lag length of a VAR?

- Even given that we know the variables to include, what lag length should we use?
  - More = better in sample fit, imprecise coefficients
  - Less = worse in sample fit, more precise coefficients, more precise forecasts.

- Lag exclusion tests
  - F tests of the exclusion of "Furthest lag" from **all** equations

- Lag Length Criteria
  - Compare alternatives with measures that penalizing larger models
    - AIC, SBC, Hannan-Quin, Maximized Likelihood,
  - Or based on forecast performance

- Excluding arbitrary lags from **individual** equations
  - Only if VERY good economic reasons, say if all subsamples have an insignificant coefficient
  - I haven't encountered such "good reasons" yet

Estimating a VAR and SVAR

1. Test for Stationarity
2. Estimate Unrestricted VAR
    1. Test for Block Exogeneity/Granger Causality

3. and investigate lag length
4. Estimate SVAR
    1. Test for congruency, parsimony
    2. Test economic hypotheses

Conclusion: what do we do?

- My opinion:
  - Best identification strategy depends on the economic story you want to tell
  - There is no short-cut, no all-convincing strategy
  - Use the one that fits the question you are trying to pose to the data
  - Do LOTS of robustness checks
    - Be honest about the outcomes
    - Argue for your preferred interpretation but acknowledge alternatives