Analysis of Multiple Time Series

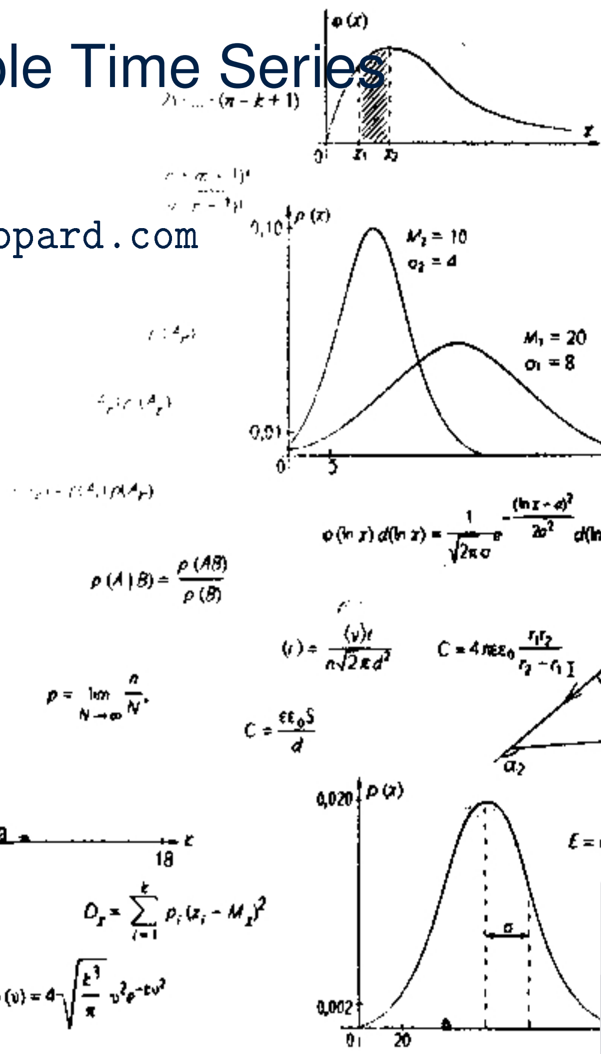
Kevin Sheppard

http://www.kevinsheppard.com

Oxford MFE

This version: February 24, 2020

February – March, 2020



DEPARTMENT OF ECONOMICS

UNIVERSITY OF OXFORD

This week's material

- Vector Autoregressions
- Basic examples
- Properties
  - Stationarity
- Revisiting univariate ARMA processes
- Forecasting
  - Granger Causality
  - Impulse Response functions
- Cointegration
  - Examining long-run relationships
  - Determining whether a VAR is cointegrated
  - Error Correction Models
  - Testing for Cointegration
    - Engle-Granger

Lots of revisiting univariate time series.

Why VAR analysis?

- Stationary VARs
  - Determine whether variables feedback into one another
  - Improve forecasts
  - Model the effect of a shock in one series on another
  - Differentiate between short-run and long-run dynamics
- Cointegration
  - Link random walks
  - Uncover long run relationships
  - Can improve medium to long term forecasting a lot

VAR Defined

- P^th order autoregression, AR(P):

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_P y_{t-p} + \epsilon_t$$

- P^th order vector autoregression, VAR(P):

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \ldots + \mathbf{\Phi}_P \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t$$

  where $\mathbf{y}_t$ and $\boldsymbol{\epsilon}_t$ are k by 1 vectors
- Bivariate VAR(1):

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- Compactly expresses two linked models:

$$y_{1,t} = \phi_{01} + \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \epsilon_{1,t}$$

$$y_{2,t} = \phi_{02} + \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \epsilon_{2,t}$$

Stationarity Revisited

- Stationarity is a statistically meaningful form of regularity. A stochastic process $\{y_t\}$ is covariance stationary if

$$
\begin{aligned}
\mathrm{E}[y_t] &= \mu && \forall t \\
\mathrm{V}[y_t] &= \sigma^2 && \sigma^2 < \infty \forall t \\
\mathrm{E}[(y_t - \mu)(y_{t-s} - \mu)] &= \gamma_s && \forall t, s
\end{aligned}
$$

- AR(1) stationarity: $y_t = \phi y_{t-1} + \epsilon_t$
  - $|\phi| < 1$
  - ϵ_t is white noise
- AR(P) stationarity: $y_t = \phi_1 y_{t-1} + \ldots + \phi_P y_{t-P} + \epsilon_t$
  - Roots of $(z^P - \phi_1 z^{P-1} - \phi_2 z^{P-2} - \ldots - \phi_{P-1} z - \phi_P)$ less than 1
  - ϵ_t is white noise
- No dependence on t

Relationship to AR

- AR(1)

$$
\begin{aligned}
y_t &= \phi_0 + \phi_1 y_{t-1} + \epsilon_t \\
&= \phi_0 + \phi_1(\phi_0 + \phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \phi_0 + \phi_1\phi_0 + \phi_1^2 y_{t-2} + \phi_1\epsilon_{t-1} + \epsilon_t \\
&= \phi_0 + \phi_1\phi_0 + \phi_1^2(\phi_0 + \phi_1 y_{t-3} + \epsilon_{t-2}) + \phi_1\epsilon_{t-1} + \epsilon_t \\
&= \phi_0 \sum_{i=0}^{\infty} \phi_1^i + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i} \\
&= (1-\phi_1)^{-1}\phi_0 + \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}
\end{aligned}
$$

- VAR(1)

$$
\begin{aligned}
\mathbf{y}_t &= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 (\mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-2} + \boldsymbol{\epsilon}_{t-1}) + \boldsymbol{\epsilon}_t \\
&= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{\Phi}_0 + \mathbf{\Phi}_1^2 \mathbf{y}_{t-2} + \mathbf{\Phi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{\Phi}_0 + \mathbf{\Phi}_1^2 (\mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-3} + \boldsymbol{\epsilon}_{t-2}) + \mathbf{\Phi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \mathbf{\Phi}_0 + \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i} \\
&= (\mathbf{I}_k - \mathbf{\Phi}_1)^{-1} \mathbf{\Phi}_0 + \sum_{i=0}^{\infty} \mathbf{\Phi}_1^i \boldsymbol{\epsilon}_{t-i}
\end{aligned}
$$

Properties of a VAR(1) and AR(1)

$$\text{AR(1)} : y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$
$$\text{VAR(1)} : \mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

	AR(1)	VAR(1)
Mean	$\phi_0/(1-\phi_1)$	$(\mathbf{I}_k - \mathbf{\Phi}_1)^{-1}\mathbf{\Phi}_0$
Variance	$\sigma^2/(1-\phi_1^2)$	$(\mathbf{I} - \mathbf{\Phi}_1 \otimes \mathbf{\Phi}_1)^{-1}\text{vec}(\mathbf{\Sigma})$
s$^{\text{th}}$ Autocovariance	$\gamma_s = \phi_1^s \mathrm{V}[y_t]$	$\mathbf{\Gamma}_s = \mathbf{\Phi}_1^s \mathrm{V}[\mathbf{y}_t]$
-s$^{\text{th}}$ Autocovariance	$\gamma_{-s} = \phi_1^s \mathrm{V}[y_t]$	$\mathbf{\Gamma}_{-s} = \mathrm{V}[\mathbf{y}_t]\mathbf{\Phi}_1^{s\prime}$

Autocovariances of vector processes are not symmetric, but $\mathbf{\Gamma}_s = \mathbf{\Gamma}'_{-s}$

- Stationarity
  - AR(1): $|\phi_1| < 1$
  - VAR(1): $|\lambda_i| < 1$ where λ_i are the eigenvalues of $\mathbf{\Phi}_1$

Stock and Bond VAR

- VWM from CRSP
- TERM constructed from 10-year bond *return minus 1-year return* from FRED
- February 1962 until December 2018 (683 months)

$$\begin{bmatrix} VWM_t \\ TERM_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ TERM_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- Market model:

$$VWM_t = \phi_{01} + \phi_{11,1} VWM_{t-1} + \phi_{12,1} 10YR_{t-1} + \epsilon_{1,t}$$

- Long bond model

$$TERM_t = \phi_{01} + \phi_{21,1} VWM_{t-1} + \phi_{22,1} TERM_{t-1} + \epsilon_{2,t}.$$

- Estimates

$$\begin{bmatrix} VWM_t \\ TERM_t \end{bmatrix} = \begin{bmatrix} \underset{(0.000)}{0.801} \\ \underset{(0.041)}{0.232} \end{bmatrix} + \begin{bmatrix} \underset{(0.122)}{0.059} & \underset{(0.004)}{0.166} \\ \underset{(0.000)}{-0.104} & \underset{(0.002)}{0.116} \end{bmatrix} \begin{bmatrix} VWM_{t-1} \\ TERM_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Stock and Bond VAR

- Estimates from VAR

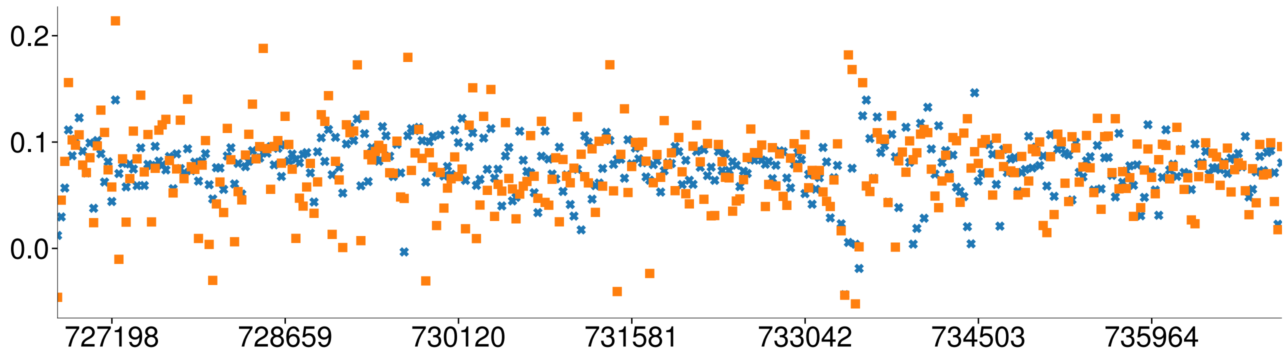
$$
\begin{aligned}
VWM_t &= \underset{(0.000)}{0.816} + \underset{(0.117)}{0.060}\ VWM_{t-1} + \underset{(0.003)}{0.168}\ TERM_{t-1} \\
TERM_t &= \underset{(0.045)}{0.228} - \underset{(0.000)}{0.104}\ VWM_{t-1} + \underset{(0.002)}{0.115}\ TERM_{t-1}
\end{aligned}
$$

- Estimates from AR

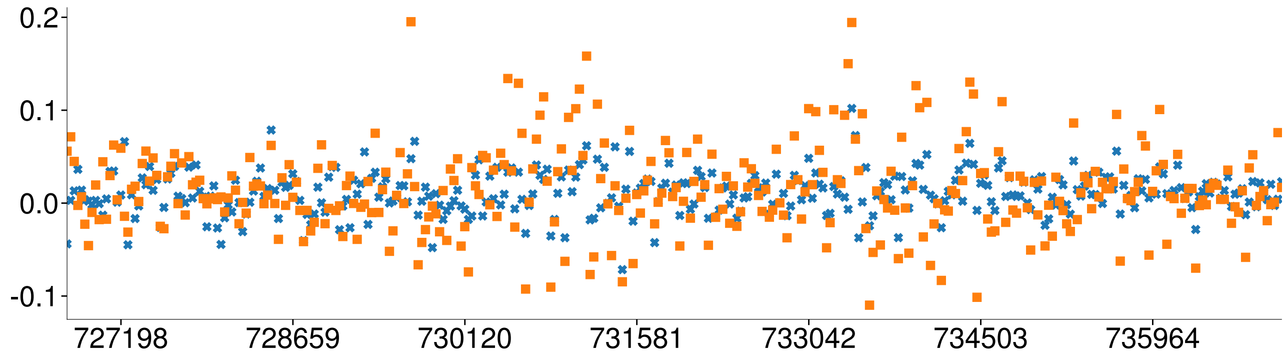
$$
\begin{aligned}
VWM_t &= \underset{(0.000)}{0.830} + \underset{(0.057)}{0.073}\ VWM_{t-1} \\
TERM_t &= \underset{(0.224)}{0.137} + \underset{(0.011)}{0.098}\ TERM_{t-1}
\end{aligned}
$$

Comparing AR and VAR forecasts

1-month-ahead forecasts of the VWM returns



1-month-ahead forecasts of 10-year bond returns



Monetary Policy VAR

- Standard tool in monetary policy analysis
  - Unemployment rate (differenced)
  - Federal Funds rate
  - Inflation rate (differenced)

$$\begin{bmatrix} \Delta\text{UNEMP}_t \\ \text{FF}_t \\ \Delta\text{INF}_t \end{bmatrix} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \begin{bmatrix} \Delta\text{UNEMP}_{t-1} \\ \text{FF}_{t-1} \\ \Delta\text{INF}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

	$\Delta \ln \text{UNEMP}_{t-1}$	FF_{t-1}	ΔINF_{t-1}
$\Delta \ln \text{UNEMP}_t$	0.624 (0.000)	0.015 (0.001)	0.016 (0.267)
FF_t	-0.816 (0.000)	0.979 (0.000)	-0.045 (0.317)
ΔINF_t	-0.501 (0.010)	-0.009 (0.626)	-0.401 (0.000)

Interpreting Estimates

- Variable scale affects cross-parameter estimates
  - Not an issue in ARMA analysis
- Standardizing data can improve interpretation when scales differ

	$\Delta \ln \mathrm{UNEMP}_{t-1}$	FF_{t-1}	$\Delta\mathrm{INF}_{t-1}$
$\Delta \ln \mathrm{UNEMP}_t$	0.624 (0.000)	0.153 (0.001)	0.053 (0.267)
FF_t	−0.080 (0.000)	0.979 (0.000)	−0.015 (0.317)
$\Delta\mathrm{INF}_t$	−0.151 (0.010)	−0.028 (0.626)	−0.401 (0.000)

- Other important measures – statistical significance, persistence, model selection – are unaffected by standardization

VAR(P) is really a VAR(1)

- Companion form:

$$\mathbf{y}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \boldsymbol{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

- Reform into a single VAR(1) where $\boldsymbol{\mu} = \mathrm{E}\left[\mathbf{y}_t\right] = \left(\mathbf{I} - \boldsymbol{\Phi}_1 - \ldots - \boldsymbol{\Phi}_P\right)^{-1} \boldsymbol{\Phi}_0$

$$\mathbf{z}_t = \boldsymbol{\Upsilon} \mathbf{z}_{t-1} + \boldsymbol{\xi}_t$$

$$\mathbf{z}_t = \begin{bmatrix} \mathbf{y}_t - \boldsymbol{\mu} \\ \mathbf{y}_{t-1} - \boldsymbol{\mu} \\ \vdots \\ \mathbf{y}_{t-P+1} - \boldsymbol{\mu} \end{bmatrix}, \quad \boldsymbol{\Upsilon} = \begin{bmatrix} \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \boldsymbol{\Phi}_3 & \ldots & \boldsymbol{\Phi}_{P-1} & \boldsymbol{\Phi}_P \\ \mathbf{I}_k & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_k & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{I}_k & \mathbf{0} \end{bmatrix}$$

  - All results can be directly applied to the companion form.
  - Can also be used to transform AR(P) into VAR(1)

Revisiting Univariate Forecasting

- Consider standard AR(1)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$

- Optimal 1-step ahead forecast:

$$\begin{aligned} \mathrm{E}_t[y_{t+1}] &= \mathrm{E}_t[\phi_0] + \mathrm{E}_t[\phi_1 y_t] + \mathrm{E}_t[\epsilon_{t+1}] \\ &= \phi_0 + \phi_1 y_t + 0 \end{aligned}$$

- Optimal 2-step ahead forecast:

$$\begin{aligned} \mathrm{E}_t[y_{t+2}] &= \mathrm{E}_t[\phi_0] + \mathrm{E}_t[\phi_1 y_{t+1}] + \mathrm{E}_t[\epsilon_{t+2}] \\ &= \phi_0 + \phi_1 \mathrm{E}_t[y_{t+1}] + 0 \\ &= \phi_0 + \phi_1(\phi_0 + \phi_1 y_t) \\ &= \phi_0 + \phi_1 \phi_0 + \phi_1^2 y_t \end{aligned}$$

- Optimal h-step ahead forecast:

$$\mathrm{E}_t[y_{t+h}] = \sum_{i=0}^{h-1} \phi_1^i \phi_0 + \phi_1^h y_t$$

Forecasting with VARs

- Identical to univariate case

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

- Optimal 1-step ahead forecast:

$$\begin{aligned} \mathrm{E}_t[\mathbf{y}_{t+1}] &= \mathrm{E}_t[\mathbf{\Phi}_0] + \mathrm{E}_t[\mathbf{\Phi}_1 \mathbf{y}_t] + \mathrm{E}_t[\boldsymbol{\epsilon}_{t+1}] \\ &= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_t + \mathbf{0} \end{aligned}$$

- Optimal h-step ahead forecast:

$$\begin{aligned} \mathrm{E}_t[y_{t+h}] &= \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{\Phi}_0 + \ldots + \mathbf{\Phi}_1^{h-1} \mathbf{\Phi}_0 + \mathbf{\Phi}_1^h \mathbf{y}_t \\ &= \sum_{i=0}^{h-1} \mathbf{\Phi}_1^i \mathbf{\Phi}_0 + \mathbf{\Phi}_1^h \mathbf{y}_t \end{aligned}$$

- Higher order forecast can be recursively computed

$$\mathrm{E}_t[\mathbf{y}_{t+h}] \quad = \quad \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathrm{E}_t[\mathbf{y}_{t+h-1}] + \ldots + \mathbf{\Phi}_P \mathrm{E}_t[\mathbf{y}_{t+h-P}]$$

What makes a good forecast?

- Forecast residuals

$$\hat{e}_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$$

- Residuals are *not* white noise
- Can contain an MA($h-1$) component
  - Forecast error for $y_{t+1} - \hat{y}_{t+1|t-h+1}$ was not known at time t.
- Plot your residuals
- Residual ACF
- Mincer-Zarnowitz regressions
- Three period procedure
  - Training sample: Used to build model
  - Validation sample: Used to refine model
  - Evaluation sample: Ultimate test, ideally 1 shot

Multi-step Forecasting

- Two methods
- Iterative method
  - Build model for 1-step ahead forecasts

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

  - Iterate forecast out to period h

$$\hat{\mathbf{y}}_{t+h|t} = \sum_{i=0}^{h-1} \mathbf{\Phi}_1^i \mathbf{\Phi}_0 + \mathbf{\Phi}_1^h \mathbf{y}_t$$

  - Makes efficient use of information
  - Imposes a lot of structure on the problem
- Direct Method
  - Build model for h-step ahead forecasts

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_h \mathbf{y}_{t-h} + \boldsymbol{\epsilon}_t$$

  - Directly forecast using a pseudo 1-step ahead method

$$\hat{\mathbf{y}}_{t+h|t} = \mathbf{\Phi}_0 + \mathbf{\Phi}_h \mathbf{y}_t$$

  - Robust to some nonlinearities

Multi-step Forecast Evaluation

- Multistep forecast evaluation is identical to one-step ahead forecast evaluation with one caveat
- h-step ahead forecast errors may be correlated with any forecast error not known at time t

$$\hat{e}_{t+1|t-h+1}, \hat{e}_{t+2|t-h+2}, \ldots, \hat{e}_{t+h-1|t-1}$$

- Leads to a MA($h-1$) structure in the forecast errors
- Solutions:
  - Use regular GMZ regression with a Newey-West covariance estimator
  $$y_{t+h} - \hat{y}_{t+h|t} = \beta_1 + \beta_2 \hat{y}_{t+h|t} + \boldsymbol{\gamma} \mathbf{x}_t + \eta_t$$
  $H_0 : \beta_1 = \beta_2 = \boldsymbol{\gamma} = 0$, $H_1 : \beta_1 \neq 0 \cup \beta_2 \neq 0 \cup \gamma_j \neq 0 \; \exists j$
  - Explicitly model the MA($h-1$) and use a standard covariance estimator
  $$y_{t+h} - \hat{y}_{t+h|t} = \beta_1 + \beta_2 \hat{y}_{t+h|t} + \boldsymbol{\gamma} \mathbf{x}_t + \eta_t + \sum_{i=1}^{h-1} \theta_i \eta_{t-i}$$
  Note: Null is the same; does not impose a restriction on $\boldsymbol{\theta}$

Example: Monetary Policy VAR

- Forecasts produced iteratively for 1 to 8 quarters ahead
- Random walk (FF) or constant mean benchmark
- AR and VAR select lag length using BIC
- Restricted force reversion to in-sample mean using 2-step estimator
  1. Estimate sample mean, and subtract to produce $\tilde{\mathbf{y}}_t = \mathbf{y}_t - \hat{\boldsymbol{\mu}}$
  2. Estimate VAR *without* a constant

$$\tilde{\mathbf{y}}_t = \mathbf{\Phi}_1 \tilde{\mathbf{y}}_{t-1} + \ldots + \mathbf{\Phi}_P \tilde{\mathbf{y}}_{t-P} + \boldsymbol{\epsilon}_t$$

  3. Forecast and then add the in-sample mean

$$\mathrm{E}_t \left[\tilde{\mathbf{y}}_{t+h} \right] + \hat{\boldsymbol{\mu}}$$

- Evaluation based on relative MSE

$$\text{Rel. MSE} = \frac{\text{MSE}}{\text{MSE}_{bm}}, \ \text{MSE} = {}^1\!/_{T-h-R} \sum_{t=R}^{T-h} \left(y_{t+h} - \hat{y}_{t+h|t} \right)^2$$

Example: Monetary Policy VAR

Horizon	Series	VAR Restricted	VAR Unrestricted	AR Restricted	AR Unrestricted
1	Unemployment	0.522	0.520	**0.507**	0.507
	Fed. Funds Rate	**0.887**	0.903	0.923	0.933
	Inflation	0.869	0.868	**0.839**	0.840
2	Unemployment	0.716	**0.710**	0.717	0.718
	Fed. Funds Rate	**0.923**	0.943	*1.112*	*1.130*
	Inflation	*1.082*	*1.081*	*1.031*	*1.030*
4	Unemployment	0.872	**0.861**	0.937	0.940
	Fed. Funds Rate	**0.952**	0.976	*1.082*	*1.109*
	Inflation	*1.000*	0.999	0.998	**0.998**
8	Unemployment	0.820	**0.806**	0.973	0.979
	Fed. Funds Rate	**0.974**	*1.007*	*1.062*	*1.110*
	Inflation	*1.001*	1.000	0.998	**0.997**

Estimation and Identification

- Univariate Identification: Box-Jenkins
  - Use ACF and PACF to determine AR and MA lag order
  - Examine residuals
  - Parsimony principle
- The autocorrelation of a scalar process is defined

$$\rho_s = \frac{\gamma_s}{\gamma_0}$$

  where γ_s is s$^{\text{th}}$ the autocovariance
  - Regression coefficient:

$$y_t = \mu + \rho_s y_{t-s} + \epsilon_t$$

- Partial autocorrelation ψ_s
  - Regression interpretation of s$^{\text{th}}$ partial autocorrelation:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_{s-1} y_{t-s+1} + \psi_s y_{t-s} + \epsilon_t$$

  - ψ is the s$^{\text{th}}$ partial autocorrelation

CCF and Partial CCF

- Multivariate equivalents
  - ACF and PACF have same regression definitions
  - Cross-correlation function

$$\rho_{xy,s} = \frac{\mathrm{E}[(x_t - \mu_x)(y_{t-s} - \mu_y)]}{\sqrt{\mathrm{V}[x_t]\mathrm{V}[y_t]}}$$

$$\rho_{yx,s} = \frac{\mathrm{E}[(y_t - \mu_y)(x_{t-s} - \mu_x)]}{\sqrt{\mathrm{V}[x_t]\mathrm{V}[y_t]}}$$

  - Generally different
  - Cross-partial-correlation function $\psi_{xy,s}$

$$\begin{aligned} x_t =& \phi_0 + \phi_{x1} x_{t-1} + \ldots + \phi_{xs-1} x_{t-(s-1)} \\ &+ \phi_{y1} y_{t-1} + \ldots + \phi_{ys-1} y_{t-(s-1)} + \varphi_{xy,s} y_{t-s} + \epsilon_{x,t} \end{aligned}$$

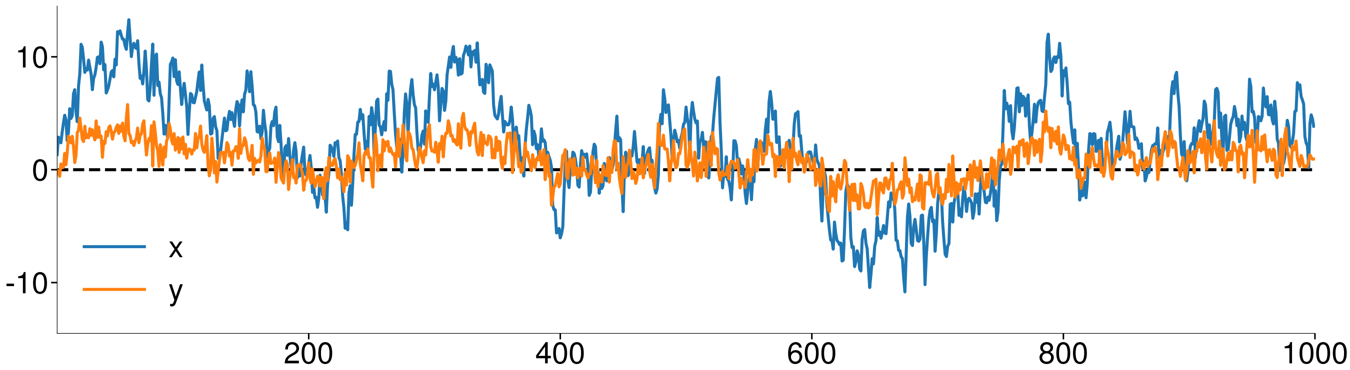
    - Can help identify VAR order
- Deeper issue: too many and too complicated
- Simple solution: Model selection

Interpreting CCFs and PCCFs

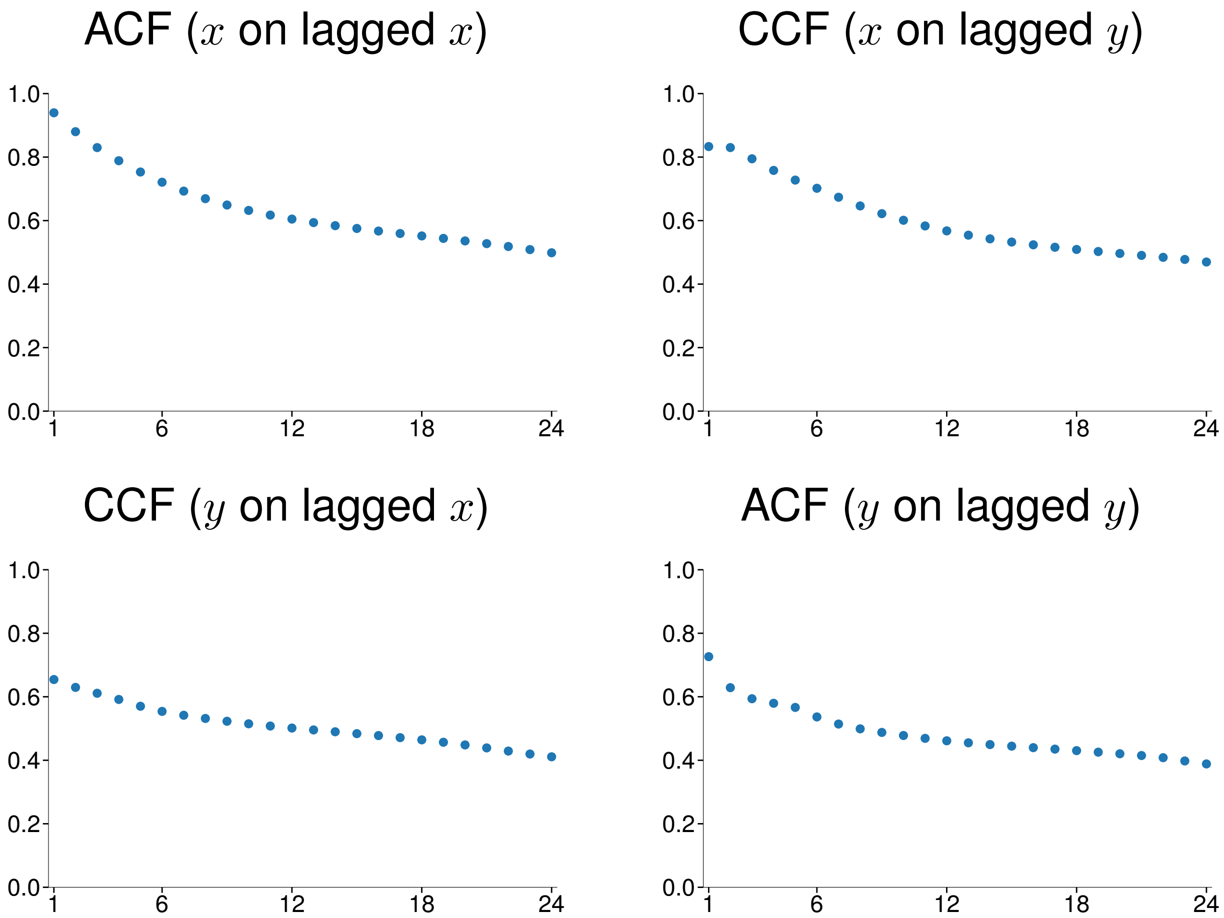
- y has HAR dynamics, spills over to x

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} 0.5 & 0.9 \\ .0 & 0.47 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \sum_{i=2}^{5} \begin{bmatrix} 0 & 0 \\ 0 & 0.06 \end{bmatrix} \begin{bmatrix} x_{t-i} \\ y_{t-i} \end{bmatrix}$$
$$+ \sum_{j=6}^{22} \begin{bmatrix} 0 & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} x_{t-j} \\ y_{t-j} \end{bmatrix} + \begin{bmatrix} \epsilon_{x,t} \\ \epsilon_{y,t} \end{bmatrix}$$

- Simulated data

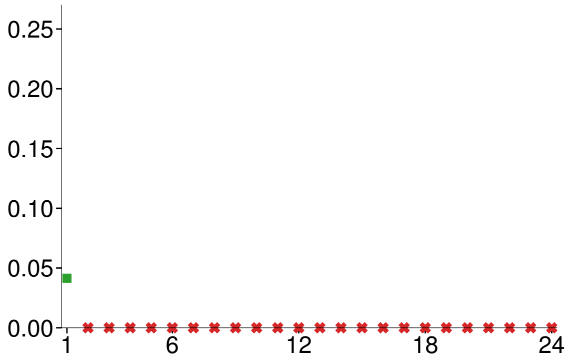


ACFs and CCFs

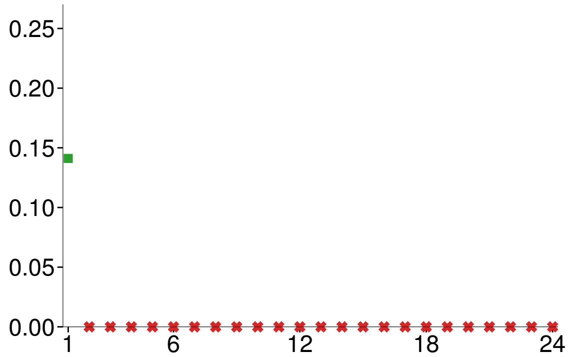


PACFs and Partial CCFs

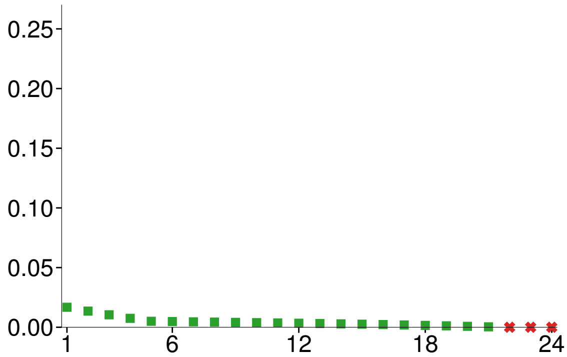
PACF (x on lagged x)



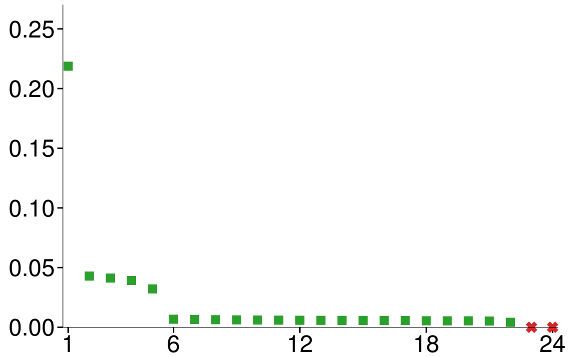
PCCF (x on lagged y)



PCCF (y on lagged x)



PACF (y on lagged y)



Model Selection

- Step 1: Pick maximum lag length
  - Information criteria

$$\begin{aligned} \text{AIC:} \quad & \ln|\mathbf{\Sigma}(P)| + k^2 P \frac{2}{T} \\ \text{Hannan-Quinn IC (HQIC):} \quad & \ln|\mathbf{\Sigma}(P)| + k^2 P \frac{\ln \ln T}{T} \\ \text{SIC:} \quad & \ln|\mathbf{\Sigma}(P)| + k^2 P \frac{\ln T}{T} \end{aligned}$$

    - $\Sigma(P)$ is the covariance of the residuals using P lags
    - $|\cdot|$ is the determinant
  - Hypothesis testing based
    - General to Specific
    - Specific to General
  - Likelihood Ratio

$$\left(T - P_2 k^2\right)\left(\ln|\mathbf{\Sigma}(P_1)| - \ln|\mathbf{\Sigma}(P_2)|\right) \stackrel{A}{\sim} \chi^2_{(P_2 - P_1)k^2}$$

Lag Length Selection in Monetary Policy VAR

- Maximum lag: 12 (1 year)

Lag Length	AIC	HQIC	BIC	LR	P-val
0	4.014	3.762	3.605	925	0.000
1	0.279	0.079	0.000▼▲	39.6	0.000
2	0.190	0.042	0.041	40.9	0.000
3	0.096	0.000▼	0.076	29.0	0.001
4	0.050▼	0.007	0.160	7.34	0.602▼
5	0.094	0.103	0.333	29.5	0.001
6	0.047	0.108	0.415	13.2	0.155
7	0.067	0.180	0.564	32.4	0.000
8	0.007	0.172▲	0.634	19.8	0.019
9	0.000▲	0.217	0.756	7.68	0.566▲
10	0.042	0.312	0.928	13.5	0.141
11	0.061	0.382	1.076	13.5	0.141
12	0.079	0.453	1.224	–	–

Granger Causality

- First fundamentally new concept
- Examines whether lags of one variable are helpful in predicting another

Definition (Granger Causality)

A scalar random variable $\{x_t\}$ is said to **not** Granger cause $\{y_t\}$ if $\mathrm{E}[y_t|x_{t-1}, y_{t-1}, x_{t-2}, y_{t-2}, \ldots] = \mathrm{E}[y_t|, y_{t-1}, y_{t-2}, \ldots]$. That is, $\{x_t\}$ does not Granger cause if the forecast of y_t is the same whether conditioned on past values of x_t or not.

Granger Causality

- Translates directly into a restriction in a VAR
- Unrestricted

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- Restricted so that x_t does not GC y_t

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$x_t = \phi_{01} + \phi_{11} x_{t-1} + \phi_{12} y_{t-1} + \epsilon_{1,t}$$

$$y_t = \phi_{02} + \phi_{22} y_{t-1} + \epsilon_{2,t} \Leftarrow \text{ No } x_t!$$

More Granger Causality

- In P lag model

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

  the null hypothesis is

$$H_0 : \phi_{ij,1} = \phi_{ij,2} = \ldots = \phi_{ij,P} = 0$$

- Alternative is

$$H_0 : \phi_{ij,1} \neq 0 \text{ or } \phi_{ij,2} \neq 0 \text{ or } \ldots \text{ or } \phi_{ij,P} \neq 0$$

- Likelihood Ratio test

$$\left(T - Pk^2\right)\left(\ln|\mathbf{\Sigma}_r| - \ln|\mathbf{\Sigma}_u|\right) \overset{A}{\sim} \chi^2_P$$

- $\mathbf{\Sigma}_u$ is the covariance of the errors from unrestricted model
- $\mathbf{\Sigma}_r$ is the covariance of the errors from restricted model
- $T - Pk^2$ is number of observations minus number of free parameters in unrestricted model
  - Why χ^2_P?

Monetary Policy VAR

- Standard tool in monetary policy analysis
  - Unemployment rate (differenced)
    - Federal Funds rate
    - Inflation rate (differenced)

$$\begin{bmatrix} \Delta\mathrm{UNEMP}_t \\ \mathrm{FF}_t \\ \Delta\mathrm{INF}_t \end{bmatrix} = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \begin{bmatrix} \Delta\mathrm{UNEMP}_{t-1} \\ \mathrm{FF}_{t-1} \\ \Delta\mathrm{INF}_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \\ \epsilon_{3,t} \end{bmatrix}.$$

Granger Causality in Campbell's VAR

- Using model with lags 3 lags (HQIC)
- $H_0 : \phi_{ij,1} = \phi_{ij,2} = \phi_{ij,3} = 0$
- $H_1 : \phi_{ij,1} \neq 0$ or $\phi_{ij,2} \neq 0$ or $\phi_{ij,3} \neq 0$
- i represent series being affected by lags of series j

	Fed. Funds Rate		Inflation		Unemployment	
Exclusion	P-val	Stat	P-val	Stat	P-val	Stat
Fed. Funds Rate	–	–	0.001	13.068	0.014	8.560
Inflation	0.001	14.756	–	–	0.375	1.963
Unemployment	0.000	19.586	0.775	0.509	–	–
All	0.000	33.139	0.000	18.630	0.005	10.472

Impulse Response Functions

- Second fundamentally new concept
- Complicated dynamics of a VAR make direct interpretation of coefficients difficult
- Solution is to examine impulse responses
- The impulse response function of y_i with respect to a shock in ϵ_j, for any j and i, is defined as the change in y_{it+s}, $s \geq 0$ for a unit shock in ϵ_{jt}
  - Hard to decipher
- As long as $\mathbf{y}_t$ is covariance stationarity it must have a VMA representation,

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \boldsymbol{\Xi}_1 \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\Xi}_2 \boldsymbol{\epsilon}_{t-2} + \ldots$$

- $\boldsymbol{\Xi}_j$ are the impulse responses!
- Why?
  - Directly measure the effect in period j of any shock

AR(P) and MA(∞)

- Any stationary AR(P)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_P y_{t-P} + \epsilon_t$$

  can be represented as an MA(∞)

$$y_t = \phi_0/(1 - \phi_1 - \phi_2 - \ldots - \phi_P) + \epsilon_t + \sum_{i=1}^{\infty} \theta_i \epsilon_{t-i}$$

- AR(1)

$$y_t = \phi_0 + \phi_1 y_{t-1} + \epsilon_t$$

  becomes

$$y_t = \phi_0/(1 - \phi_1) + \epsilon_t + \sum_{i=1}^{\infty} \phi_1^i \epsilon_{t-i}$$

- Stationary VAR(P) have the same relationship to VMA(∞)

$$\mathbf{y}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \ldots + \mathbf{\Phi}_P \mathbf{y}_{t-P} + \boldsymbol{\epsilon}_t$$

$$\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\epsilon}_t + \mathbf{\Xi}_1 \boldsymbol{\epsilon}_{t-1} + \mathbf{\Xi}_2 \boldsymbol{\epsilon}_{t-2} + \ldots$$

Solving IR

- Easy in VAR(1)

$$\mathbf{y}_t = (\mathbf{I}_K - \mathbf{\Phi}_1)^{-1}\mathbf{\Phi}_0 + \boldsymbol{\epsilon}_t + \mathbf{\Phi}_1\boldsymbol{\epsilon}_{t-1} + \mathbf{\Phi}_1^2\boldsymbol{\epsilon}_{t-2} + \ldots.$$

- $\mathbf{\Xi}_j = \mathbf{\Phi}_1^j$
- In the general VAR(P),

$$\mathbf{\Xi}_j = \mathbf{\Phi}_1\mathbf{\Xi}_{j-1} + \mathbf{\Phi}_2\mathbf{\Xi}_{j-2} + \ldots + \mathbf{\Phi}_P\mathbf{\Xi}_{j-P}$$

  where $\mathbf{\Xi}_0 = \mathbf{I}_k$ and $\mathbf{\Xi}_m = \mathbf{0}$ for $m < 0$.
  - In a VAR(2),

$$\mathbf{y}_t = \mathbf{\Phi}_1\mathbf{y}_{t-1} + \mathbf{\Phi}_2\mathbf{y}_{t-2} + \boldsymbol{\epsilon}_t$$

    - $\mathbf{\Xi}_0 = \mathbf{I}_k$, $\mathbf{\Xi}_1 = \mathbf{\Phi}_1$, $\mathbf{\Xi}_2 = \mathbf{\Phi}_1^2 + \mathbf{\Phi}_2$, and $\mathbf{\Xi}_3 = \mathbf{\Phi}_1^3 + \mathbf{\Phi}_1\mathbf{\Phi}_2 + \mathbf{\Phi}_2\mathbf{\Phi}_1$.
- Confidence intervals are also somewhat painful
  - Explained in notes

Considerations for Shocks

- Simple bivariate VAR(1)

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \phi_{01} \\ \phi_{02} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- How you *shock* matters
- Depends on correlation between $\epsilon_{1,t}$ and $\epsilon_{2,t}$
- 3 methods
  - Ignore correlation and just shock $\epsilon_{j,t}$ with a 1 standard deviation shock
  - Use Cholesky to factor $\mathbf{\Sigma}$ and use $\mathbf{\Sigma}^{1/2}\mathbf{e}_j$ where $\mathbf{e}_j$ is a vector of zeros with 1 in the j$^{\text{th}}$ position

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & .5 \\ .5 & 1 \end{bmatrix} \quad \mathbf{\Sigma}_C^{1/2} = \begin{bmatrix} 1 & 0 \\ .5 & .866 \end{bmatrix}$$

    - Variable order matters
  - "Generalized" impulse response that uses a projection method

Example of the different shocks

- Define the error covariance

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y\rho \\ \sigma_x\sigma_y\rho & \sigma_y^2 \end{bmatrix}$$

  - Standardized

$$\begin{bmatrix} \sigma_x \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ \sigma_y \end{bmatrix}$$

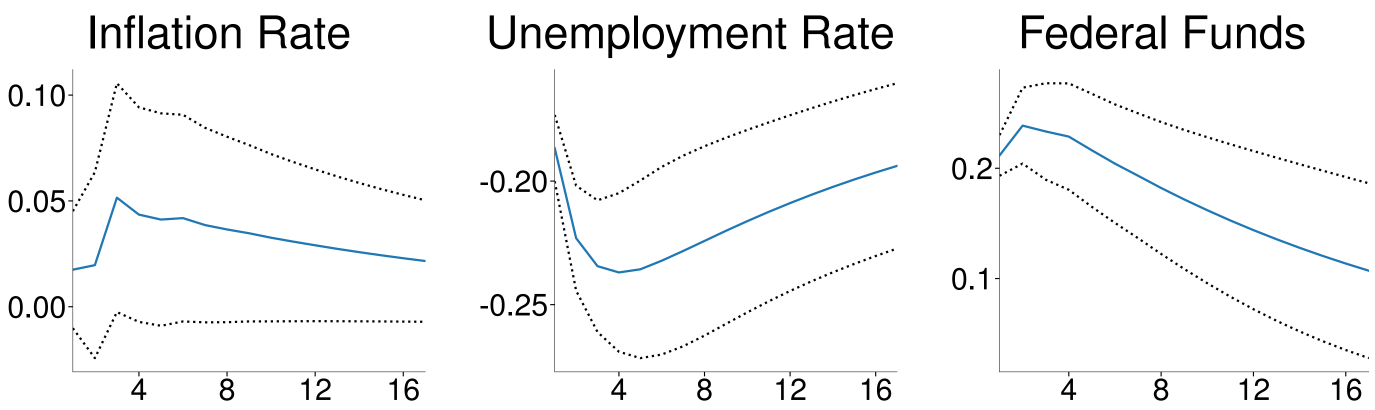
  - Cholesky

$$\mathbf{\Sigma}_C^{1/2} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \sigma_x & 0 \\ \sigma_y\rho & \sigma_y\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma_y\sqrt{1-\rho^2} \end{bmatrix}$$

$$\begin{bmatrix} \sigma_x & 0 \\ \sigma_y\rho & \sigma_y\sqrt{1-\rho^2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sigma_x \\ \sigma_y\rho \end{bmatrix} \text{,other is } \begin{bmatrix} 0 \\ \sigma_y\sqrt{1-\rho^2} \end{bmatrix}$$

Impulse Responses

- Federal Funds ordered first
- Response to Federal Funds Shock
- Cholesky factorization



Cointegration

- Cointegration is the VAR version of unit roots
- Establishes long run relationships between two unit root variables
  - Consumption has a unit root, income has a unit root
  - Consumption - Income : ????

Definition (Integrated of Order 1)

A variable y_t is integrated of order 1 (I(1)) if y_t is non-stationary and $\Delta y_t = y_t - y_{t-1}$ is stationary.

Cointegration

Definition (Bivariate Cointegration)

If x_t and y_t are are cointegrated if both are I(1) and there exists a vector $\boldsymbol{\beta}$ with both elements non-zero such that

$$\beta_1 x_t - \beta_2 y_t \sim I(0)$$

- Strong link between x_t and y_t
- Both are random walks but difference is mean reverting
- Mean reversion to the trend (stochastic trend)

What does cointegration look like?

$$\mathbf{y}_t = \mathbf{\Phi}_{ij}\mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\mathbf{\Phi}_{11} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \qquad \mathbf{\Phi}_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

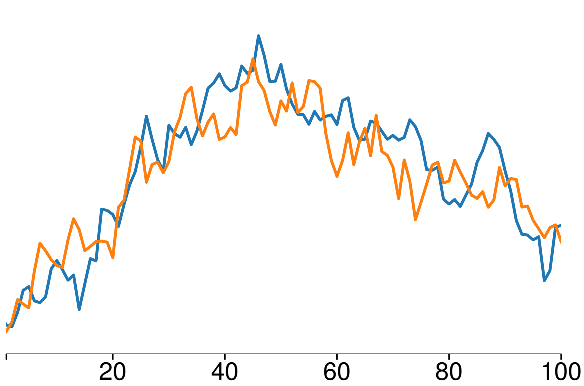
$$\lambda_i = 1, 0.6 \qquad \lambda_i = 1, 1$$

$$\mathbf{\Phi}_{21} = \begin{bmatrix} .7 & .2 \\ .2 & .7 \end{bmatrix} \qquad \mathbf{\Phi}_{22} = \begin{bmatrix} -.3 & .3 \\ .1 & -.2 \end{bmatrix}$$

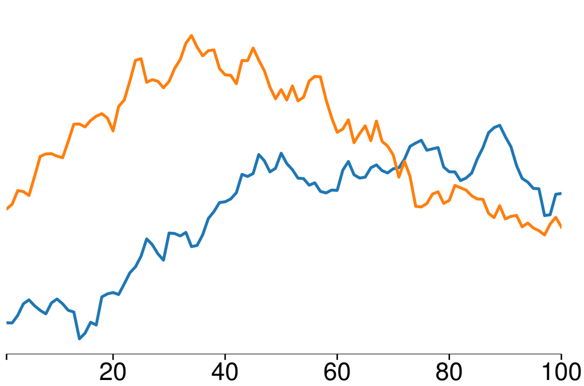
$$\lambda_i = 0.9, 0.5 \qquad \lambda_i = -0.43, -0.06$$

Persistence, Anti-persistence and Cointegration

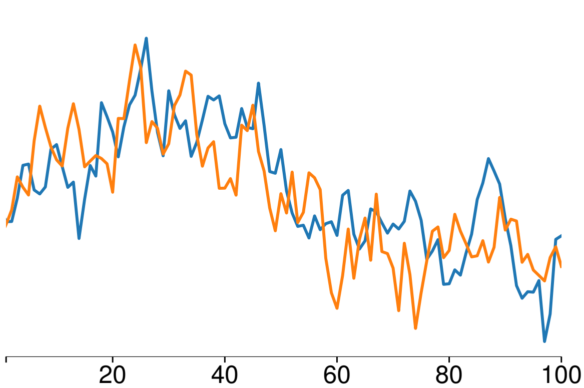
Cointegration $(\mathbf{\Phi}_{11})$



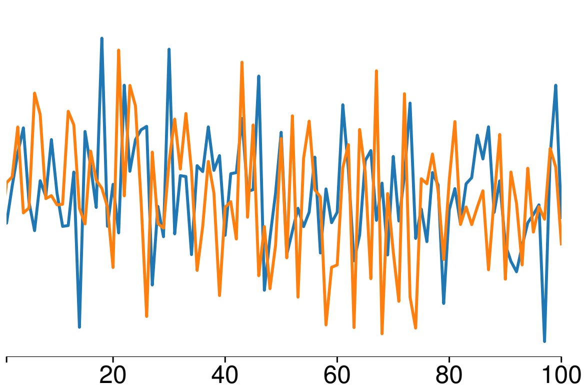
Independent Unit Roots$(\mathbf{\Phi}_{12})$



Persistent, Stationary $(\mathbf{\Phi}_{21})$



Anti-persistent, Stationary$(\mathbf{\Phi}_{22})$



How do we know when a VAR is cointegrated?

- Eigenvalue condition determines whether a VAR(1) is cointegrated

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- Cointegrated if only 1 eigenvalue is unity.
- If all less than 1: ?
- If both 1: two independent unit roots

$$\mathbf{\Phi}_{11} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \qquad \mathbf{\Phi}_{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\lambda_i = 1, 0.6 \qquad \lambda_i = 1, 1$$

$$\mathbf{\Phi}_{21} = \begin{bmatrix} .7 & .2 \\ .2 & .7 \end{bmatrix} \qquad \mathbf{\Phi}_{22} = \begin{bmatrix} -.3 & .3 \\ .1 & -.2 \end{bmatrix}$$

$$\lambda_i = 0.9, 0.5 \qquad \lambda_i = -0.43, -0.06$$

Error Correction Models

- Major point of cointegration
  - Cointegrated $\Leftrightarrow$ Error correction model
- What is an error correction model?
  - Cointegrated VAR:
$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$
  - Error correction model:
$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$
  - Normalized form
$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 \\ .2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$
- $[1 \quad -1]$ is cointegrating vector
- $[-.2 \;\; .2]'$ measures the speed of adjustment

From VAR to VECM

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Subtracting $[y_{t-1} \; x_{t-1}]'$ from both sides

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} - \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} - \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \left(\begin{bmatrix} .8 & .2 \\ .2 & .8 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 \\ .2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

Cointegrating vectors

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 & .2 \\ .2 & -.2 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

$$\begin{bmatrix} \Delta y_t \\ \Delta x_t \end{bmatrix} = \begin{bmatrix} -.2 \\ .2 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ x_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix}$$

- Cointegrating relationship can always be decomposed

$$\Delta \mathbf{y}_t = \boldsymbol{\pi} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\pi} = \boldsymbol{\alpha} \boldsymbol{\beta}'$$

- $\boldsymbol{\alpha}$ measures the speed of convergence
- $\boldsymbol{\beta}$ contain the cointegrating vectors
- Number of cointegrating vectors is rank$(\boldsymbol{\alpha}\boldsymbol{\beta}')$

$$\boldsymbol{\alpha}\boldsymbol{\beta}' = \begin{bmatrix} 0.3 & 0.2 & -0.36 \\ 0.2 & 0.5 & -0.35 \\ -0.3 & -0.3 & 0.39 \end{bmatrix}$$

- How many?

Determining the cointegrating vectors

$$\Delta \mathbf{y}_t = \boldsymbol{\pi} \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t$$

$$\boldsymbol{\pi} = \begin{bmatrix} 0.3 & 0.2 & -0.36 \\ 0.2 & 0.5 & -0.35 \\ -0.3 & -0.3 & 0.39 \end{bmatrix}$$

- Put $\boldsymbol{\pi}$ in row echelon form

$$\text{Row Echelon Form} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

- Recall $\boldsymbol{\pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -.3 \end{bmatrix} \quad \boldsymbol{\alpha} = \begin{bmatrix} .3 & .2 \\ .2 & .5 \\ -.3 & -.3 \end{bmatrix}$$

Solving for the cointegrating vectors

$$\boldsymbol{\alpha\beta}' = \begin{bmatrix} 0.3 & 0.2 & -0.36 \\ 0.2 & 0.5 & -0.35 \\ -0.3 & -0.3 & 0.39 \end{bmatrix}$$

$$\text{Row-Echelon Form} \Rightarrow \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -0.3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \beta_1 & \beta_2 \end{bmatrix}$$

and $\boldsymbol{\alpha}$ has 6 unknown parameters. $\boldsymbol{\alpha\beta}'$ can be combined to produce

$$\boldsymbol{\pi} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{11}\beta_1 + \alpha_{12}\beta_2 \\ \alpha_{21} & \alpha_{22} & \alpha_{21}\beta_1 + \alpha_{22}\beta_2 \\ \alpha_{31} & \alpha_{32} & \alpha_{31}\beta_1 + \alpha_{32}\beta_2 \end{bmatrix}$$

Testing for Cointegration

- Two tests for cointegration
  - Engle-Granger
  - Johansen
- We will focus on Engle-Granger
  - Simple and intuitive
  - Only applicable with 1 cointegrating relationship
- Test key property of cointegration: difference is I(0)
- Most of the work is a simple OLS

$$y_t = \delta_0 + \beta x_t + \epsilon_t$$

- Rest of work is testing $\hat{\epsilon}_t$ for a unit root
- Johansen tests eigenvalues of $\boldsymbol{\pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ directly.

Engle-Granger Procedure

Algorithm (Engle-Granger Test)

1. *Begin by analyzing* x_t *and* y_t *in isolation. Both must be unit roots to consider cointegration.*
2. *Estimate the long run relationship*

$$y_t = \delta_0 + \beta x_t + \epsilon_t$$

and test $H_0 : \gamma = 0$ *against* $H_0 : \gamma < 0$ *in the ADF regression*

$$\Delta \hat{\epsilon}_t = \gamma \hat{\epsilon}_{t-1} + \delta_1 \Delta \hat{\epsilon}_{t-1} + \ldots + \delta_p \Delta \hat{\epsilon}_{t-P} + \eta_t .$$

3. *Using the estimated parameters, specify and estimate the error correction form of the relationship,*

$$\left[\begin{array}{c} \Delta x_t \\ \Delta y_t \end{array}\right] = \begin{array}{c} \pi_{01} \\ \pi_{02} \end{array} + \begin{array}{c} \alpha_1 \hat{\epsilon}_t \\ \alpha_2 \hat{\epsilon}_t \end{array} + \boldsymbol{\pi}_1 \left[\begin{array}{c} \Delta x_{t-1} \\ \Delta y_{t-1} \end{array}\right] + \ldots + \boldsymbol{\pi}_P \left[\begin{array}{c} \Delta x_{t-P} \\ \Delta y_{t-P} \end{array}\right] + \left[\begin{array}{c} \eta_{1,t} \\ \eta_{2,t} \end{array}\right]$$

4. *Assess the model*

Engle-Granger Considerations

- Deterministic terms
  - No deterministic terms: only in special circumstances

$$y_t = \beta x_t + \epsilon_t$$

  - Constant: standard case

$$y_t = \delta_0 + \beta x_t + \epsilon_t$$

  - Time trend and constant: allow different growth rates/time trends in variables

$$y_t = \delta_0 + \delta_1 t + \beta x_t + \epsilon_t$$

- Critical Values
  - Critical values depend on the deterministics in the CI regression
    - Models with more deterministics have lower (more negative) critical values
  - Critical values depend on number of RHS $I(1)$ variables
    - Larger models have lower critical values

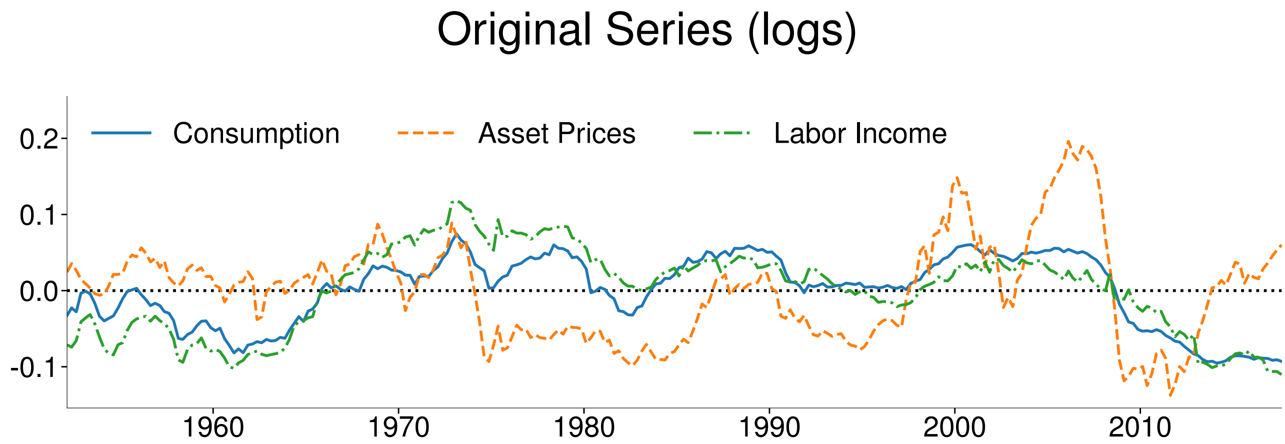
Example: cay

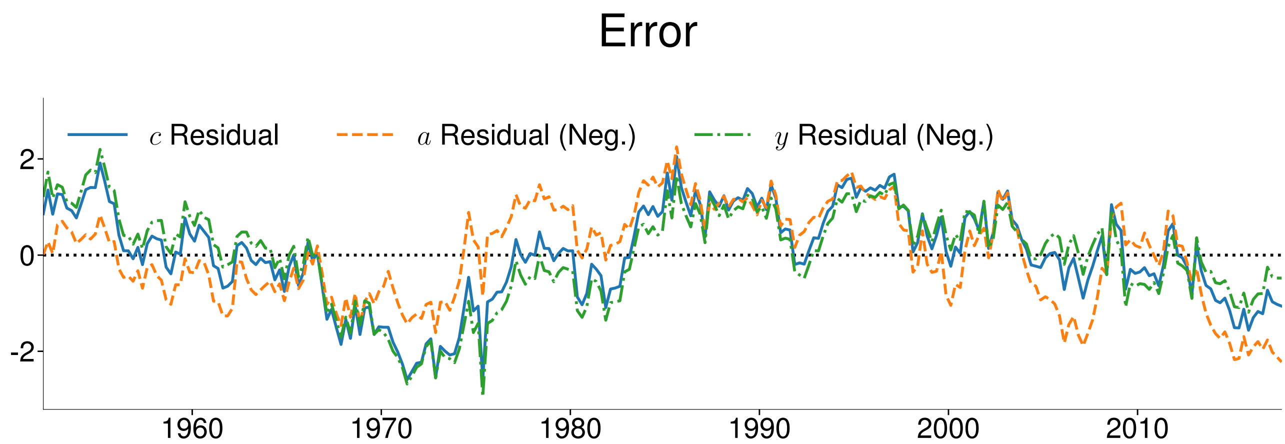
- Consumption-Aggregate Wealth has been an interesting cointegrated series in recent finance literature
- Has revived the CCAPM
- Three components:
  - Consumption (c)
  - Asset Wealth (a)
  - Labor Income (Human Wealth) (y)
- Deviation from long run related to expected return
- Cointegrating relationship: $c_t + .643 - 0.249a_t - 0.785y_t$

Unit Root Tests

Series	T-stat	P-val	ADF Lags
c	-1.198	0.674	5
a	-0.205	0.938	3
y	-2.302	0.171	0
$\hat{\epsilon}_t^c$	-2.706	0.383	1
$\hat{\epsilon}_t^a$	-2.573	0.455	0
$\hat{\epsilon}_t^y$	-2.679	0.398	1

cay Cointegration Analysis





Vector Error Correction Model

- VECM estimated using the residuals from cointegrating regression

$$\begin{bmatrix} \Delta c_t \\ \Delta a_t \\ \Delta y_t \end{bmatrix} = \begin{bmatrix} \underset{(0.000)}{0.003} \\ \underset{(0.014)}{0.004} \\ \underset{(0.000)}{0.003} \end{bmatrix} + \begin{bmatrix} \underset{(0.281)}{-0.000} \\ \underset{(0.037)}{0.002} \\ \underset{(0.515)}{0.000} \end{bmatrix} \hat{\epsilon}_{t-1} + \begin{bmatrix} \underset{(0.005)}{0.192} & \underset{(0.000)}{0.102} & \underset{(0.004)}{0.147} \\ \underset{(0.116)}{0.282} & \underset{(0.006)}{0.220} & \underset{(0.414)}{-0.149} \\ \underset{(0.000)}{0.369} & \underset{(0.088)}{0.061} & \underset{(0.140)}{-0.139} \end{bmatrix} \begin{bmatrix} \Delta c_{t-1} \\ \Delta a_{t-1} \\ \Delta y_{t-1} \end{bmatrix} + \boldsymbol{\eta}_t$$

- P-values in parentheses
- Estimation of cointegration relationship has no effect on standard errors
  - Converges fast (T)
  - VECM parameters converge at rate $\sqrt{T}$

Spurious Regression and Balance

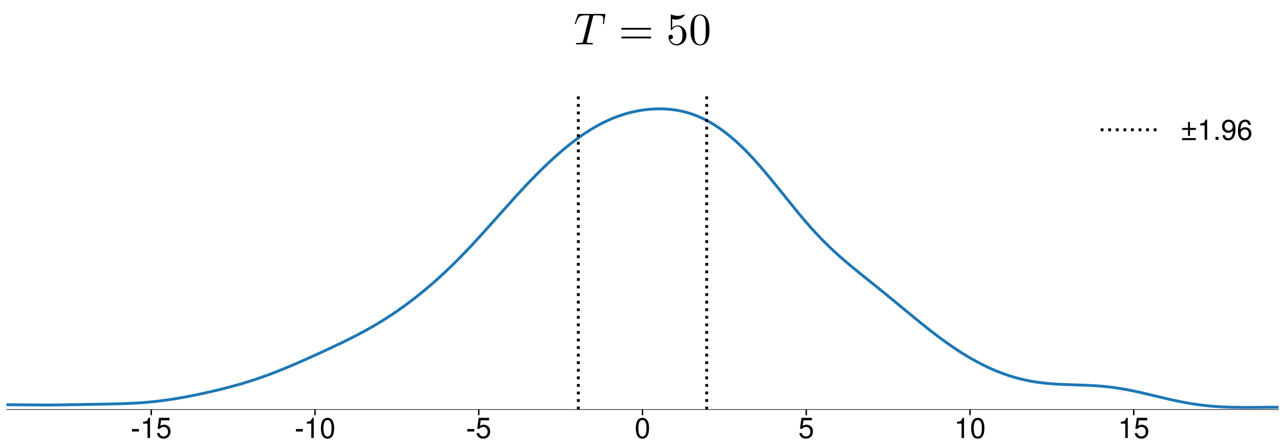
- Caution is needed when working with I(1) data
  - I(0) on I(0): The usual case. Standard asymptotic arguments apply.
  - I(1) on I(0): This regression is unbalanced.
  - I(1) on I(1): Cointegration or spurious regression.
  - I(0) on I(1): This regression is unbalanced.
- Spurious regression can lead to large t-stats when the series are independent.
  - Two unrelated I(1) processes, x_t and y_t

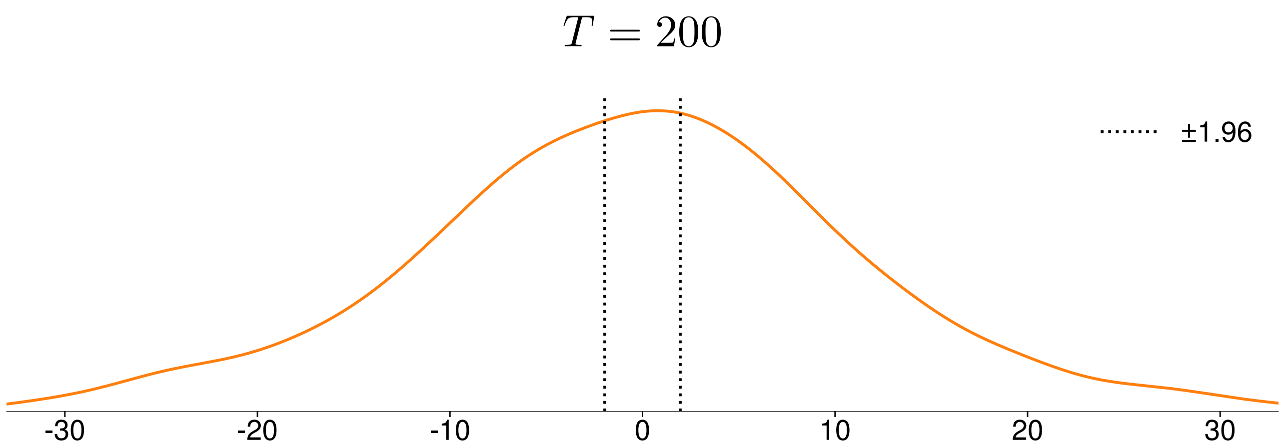
$$x_t = x_{t-1} + \epsilon_t$$

$$y_t = y_{t-1} + \eta_t$$

  - When $T = 50$, approx 80% of t-stats are significant
  - Always check for I(1) when using time-series data
  - If both I(1), make sure cointegrated.

Spurious Regression





Cross-section Regression with Time Series Data

- It is common to run regressions using time-series data

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + \epsilon_t$$

- Using time-series data in a cross-sectional regression may require modification to inference
- Modification is needed if the scores $\{\mathbf{x}_t \epsilon_t\}$ are autocorrelated

$$\begin{aligned}
\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \left(\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \epsilon_t \\
\Rightarrow \mathrm{V}\left[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right] &\approx \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1} \mathrm{V}\left[\frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t \epsilon_t \right] \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}
\end{aligned}$$

  - Usually occurs when the errors ϵ_t are autocorrelated due to mis- or under-specification of the model

Why the difference?

- Consider the estimation of the mean when y_t has white noise errors

$$y_t = \mu + \epsilon_t$$

- Obviously
- The sample mean is unbiased

$$\begin{aligned} \mathrm{E}[\hat{\mu}] &= \mathrm{E}\left[T^{-1}\sum_{t=1}^{T} y_t\right] \\ &= T^{-1}\sum_{t=1}^{T} \mathrm{E}\left[y_t\right] \\ &= \mu \end{aligned}$$

Why the difference?

- The variance of the sample mean

$$
\begin{aligned}
\mathrm{V}[\hat{\mu}] &= \mathrm{E}\left[\left(T^{-1}\sum_{t=1}^{T} y_t - \mu\right)^2\right] \\
&= \mathrm{E}\left[T^{-2}\left(\sum_{t=1}^{T}\epsilon_t^2 + \sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T}\epsilon_r\epsilon_s\right)\right] \\
&= T^{-2}\sum_{t=1}^{T}\mathrm{E}[\epsilon_t^2] + T^{-2}\sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T}\mathrm{E}[\epsilon_r\epsilon_s] \\
&= T^{-2}\sum_{t=1}^{T}\sigma^2 + T^{-2}\sum_{r=1}^{T}\sum_{s=1,r\neq s}^{T} 0 \\
&= \frac{\sigma^2}{T},
\end{aligned}
$$

- Due to white noise, $\mathrm{E}[\epsilon_i\epsilon_j]$=0 whenever $i \neq j$.
- This is the usual result

The case of an MA(1) error

- Now suppose that the error follows an MA(1)

$$\eta_t = \theta\epsilon_{t-1} + \epsilon_t$$

  where $\{\epsilon_t\}$ is a white noise process
- Error is mean 0 and so sample mean is still unbiased
- Variance of sample mean is *different* since η_t is autocorrelated
  - $\mathrm{E}[\eta_t\eta_{t-1}] \neq 0.$

$$\begin{aligned}
\mathrm{V}[\hat{\mu}] &= \mathrm{E}\left[\left(T^{-1}\sum_{t=1}^{T}\eta_t\right)^2\right] \\
&= \mathrm{E}\left[T^{-2}\left(\sum_{t=1}^{T}\eta_t^2 + 2\sum_{t=1}^{T-1}\eta_t\eta_{t+1} + 2\sum_{t=1}^{T-2}\eta_t\eta_{t+2} + \ldots + \right.\right. \\
&\qquad \left.\left. 2\sum_{t=1}^{2}\eta_t\eta_{t+T-2} + 2\sum_{t=1}^{1}\eta_t\eta_{t+T-1}\right)\right]
\end{aligned}$$

The case of an MA(1) error

- In terms of autocovariances,

$$
\begin{aligned}
\mathrm{V}[\hat{\mu}] = T^{-2}\sum_{t=1}^{T}\mathrm{E}[\eta_t^2] + 2T^{-2}\sum_{t=1}^{T-1}\mathrm{E}[\eta_t\eta_{t+1}] + 2T^{-2}\sum_{t=1}^{T-2}\mathrm{E}[\eta_t\eta_{t+2}] + \ldots + \\
2T^{-2}\sum_{t=1}^{2}\mathrm{E}[\eta_t\eta_{t+T-2}] + 2T^{-2}\sum_{t=1}^{1}\mathrm{E}[\eta_t\eta_{t+T-1}] \\
= T^{-2}\sum_{t=1}^{T}\gamma_0 + 2T^{-2}\sum_{t=1}^{T-1}\gamma_1 + 2T^{-2}\sum_{t=1}^{T-2}\gamma_2 + \ldots + 2T^{-2}\sum_{t=1}^{1}\gamma_{T-1}
\end{aligned}
$$

- $\gamma_0 = \mathrm{V}[\eta_t]$=$\left(1+\theta^2\right)\mathrm{V}\left[\epsilon_t\right]$ and $\gamma_s = \mathrm{E}[\eta_t\eta_{t-s}]$
- An MA(1) has 1 non-zero autocovariance,

$$
\begin{aligned}
\gamma_1 &= \mathrm{E}[\eta_t\eta_{t-1}] \\
&= \mathrm{E}[(\theta\epsilon_{t-1}+\epsilon_t)(\theta\epsilon_{t-2}+\epsilon_{t-1})] \\
&= \theta^2\mathrm{E}[\epsilon_{t-1}\epsilon_{t-2}] + \theta\mathrm{E}[\epsilon_{t-1}^2] + \theta\mathrm{E}[\epsilon_t\epsilon_{t-2}] + \mathrm{E}[\epsilon_t\epsilon_{t-1}] \\
&= \theta\sigma^2
\end{aligned}
$$

The case of an MA(1) error

- Putting it all together

$$
\begin{aligned}
\mathrm{V}[\hat{\mu}] &= T^{-2}\sum_{t=1}^{T}\gamma_0 + 2T^{-2}\sum_{t=1}^{T+1}\gamma_1 \\
&= T^{-2}T\gamma_0 + 2T^{-2}(T-1)\gamma_1 \\
&\approx \frac{\gamma_0 + 2\gamma_1}{T} \\
&= \frac{\sigma^2\left(1+\theta^2+2\theta\right)}{T}
\end{aligned}
$$

Can be larger or smaller ($-2 < \theta < 0$)

The variance of the sum is the sum of the variance
only when the errors are uncorrelated

Estimating the parameter covariance (from CS lecture)

- When the scores are uncorrelated (a Martingale Difference sequence (MDS)) White's covariance estimator is consistent

Theorem (Consistency of Asymptotic Covariance Estimator)

Under the large sample assumptions,

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}} &= T^{-1}\mathbf{X}'\mathbf{X} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{XX}} \\ \hat{\mathbf{S}} &= T^{-1}\sum_{t=1}^{T} \hat{\epsilon}_t^2 \mathbf{x}_t' \mathbf{x}_t \xrightarrow{p} \mathbf{S}\end{aligned}$$

and

$$\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{-1}\hat{\mathbf{S}}\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}^{-1} \xrightarrow{p} \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}\mathbf{S}\boldsymbol{\Sigma}_{\mathbf{XX}}^{-1}$$

Modification to regression parameter covariance

- White's estimator is only heteroskedasticity robust – not heteroskedasticity and autocorrelation robust

$$\hat{\mathbf{S}} = T^{-1}\sum_{t=1}^{T}\hat{\epsilon}_t^2\mathbf{x}_t'\mathbf{x}_t \xrightarrow{p} \mathbf{S}$$

- Solution is to use a Newey-West covariance for the scores ($\mathbf{x}_t\epsilon_t$)

Definition (Newey-West Covariance Estimator)

Let $\mathbf{z}_t$ be a k by 1 vector series that may be autocorrelated and define $\mathbf{z}_t^* = \mathbf{z}_t - \bar{\mathbf{z}}$ where $\bar{\mathbf{z}} = T^{-1}\sum_{t=1}^{T}\mathbf{z}_t$. The L-lag Newey-West covariance estimator for the variance of $\bar{\mathbf{z}}$ is

$$\hat{\mathbf{\Sigma}}_{NW} = \hat{\mathbf{\Gamma}}_0 + \sum_{l=1}^{L} w_l\left(\hat{\mathbf{\Gamma}}_l + \hat{\mathbf{\Gamma}}_l'\right)$$

where $\hat{\mathbf{\Gamma}}_l = T^{-1}\sum_{t=l+1}^{T}\mathbf{z}_t^*\mathbf{z}_{t-l}^{*\prime}$ and $w_l = 1 - \frac{l}{L+1}$.

Modification to regression parameter covariance

- Applied to a cross-sectional regression with time-series data

$$\hat{\mathbf{S}}_{NW} = T^{-1}\left(\sum_{t=1}^{T} e_t^2 \mathbf{x}_t'\mathbf{x}_t + \sum_{l=1}^{L} w_l \left(\sum_{s=l+1}^{T} e_s e_{s-l}\mathbf{x}_s'\mathbf{x}_{s-l} + \sum_{q=l+1}^{T} e_{q-l} e_q \mathbf{x}_{q-l}'\mathbf{x}_q\right)\right)$$
$$= \hat{\mathbf{\Gamma}}_0 + \sum_{l=1}^{L} w_l \left(\hat{\mathbf{\Gamma}}_l + \hat{\mathbf{\Gamma}}_l'\right)$$

- The HAC robust covariance of $\hat{\beta}$ is

$$\hat{\mathbf{\Sigma}}_{\mathbf{XX}}^{-1}\hat{\mathbf{S}}_{NW}\hat{\mathbf{\Sigma}}_{\mathbf{XX}}^{-1}$$

Considerations when using Newey-West an estimator

- Is a Newey-West estimator needed? Complex estimators have worse finite sample performance
- It **must** be the case that $L \to \infty$ as $T \to \infty$
- Even if the scores follow a MA(1)!
- Optimal rate is $O(T^{\frac{1}{3}})$ so $L \propto T^{\frac{1}{3}}$ or $L = cT^{\frac{1}{3}}$ for some (unknown) c
- Other HAC estimators available and may work well if the scores very persistent
  - Den Haan-Levin
- Alternative is to include lagged regressand(s) in the regression

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + \sum_{p=1}^{P} \phi_p y_{t-p} + \epsilon_t$$

  - Not popular when focus is on cross-section component of model