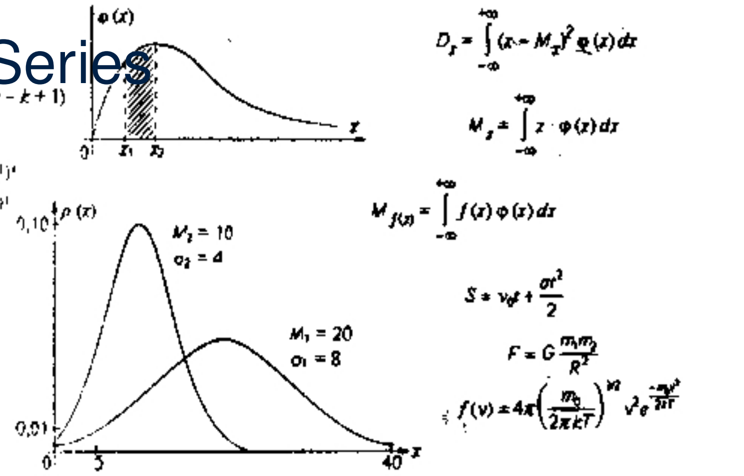
Analysis of Multiple Time Series

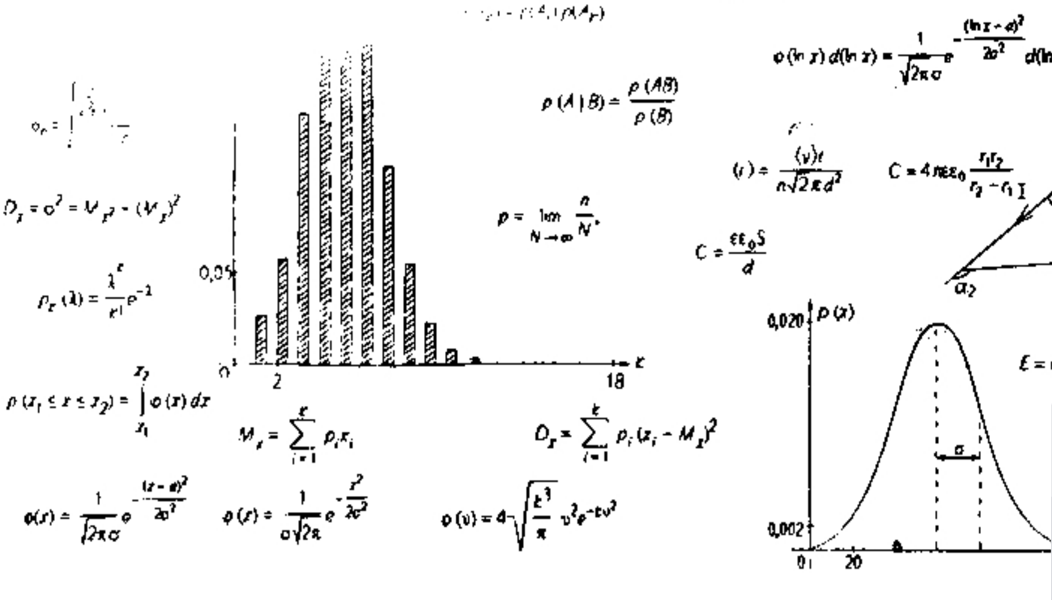
Kevin Sheppard
http://www.kevinsheppard.com

Oxford MFE

This version: February 25, 2019

February 21, 2019





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.

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^{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

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

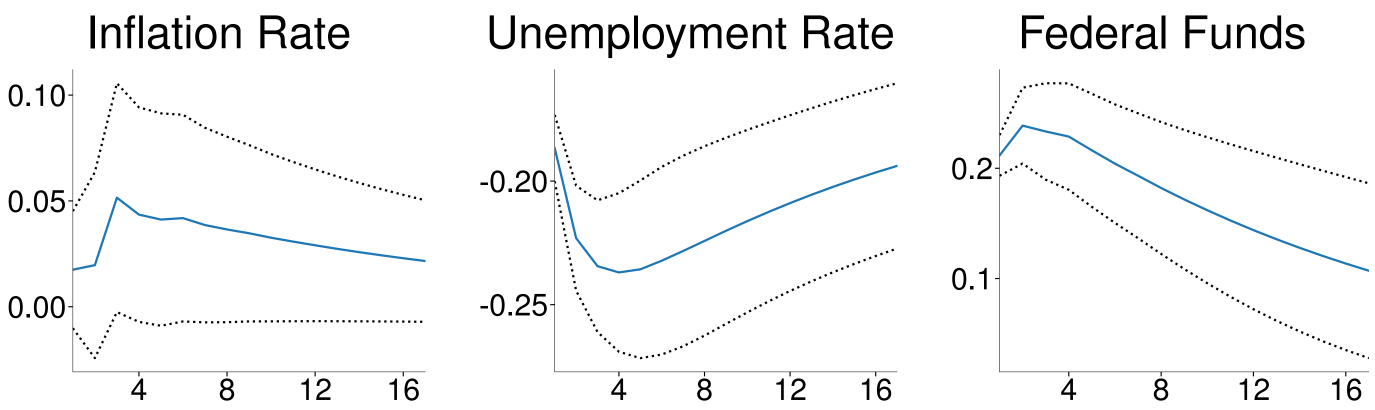
  - Cholesky

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

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

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?

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

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

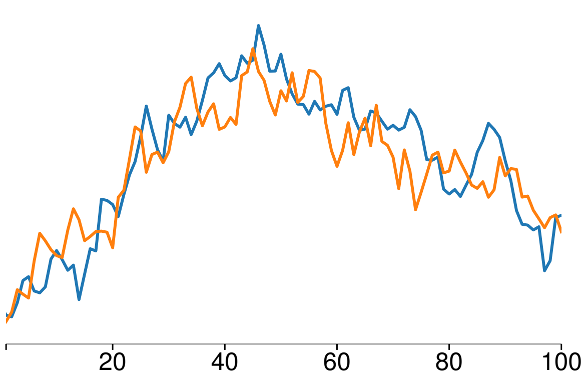
$$\lambda_i = 1, 0.6 \qquad \lambda_i = 1, 1$$

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

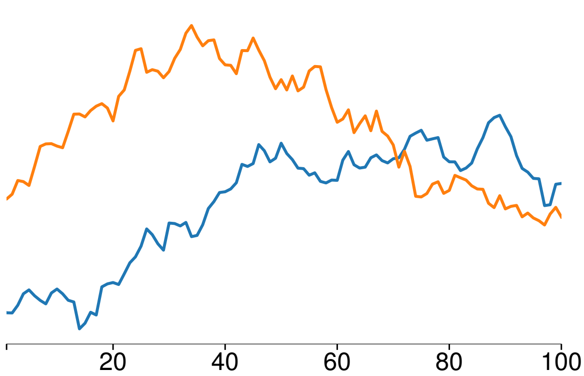
$$\lambda_i = 0.9, 0.5 \qquad \lambda_i = -0.43, -0.06$$

Persistence, Anti-persistence and Cointegration

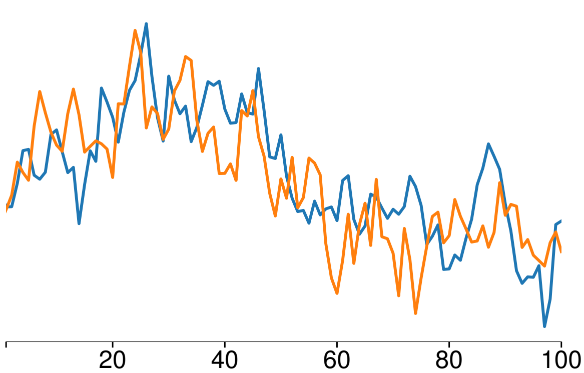
Cointegration $(\mathbf{\Phi}_{11})$



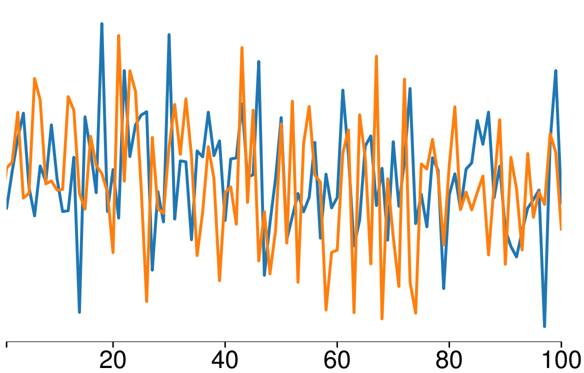
Independent Unit Roots$(\mathbf{\Phi}_{12})$



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

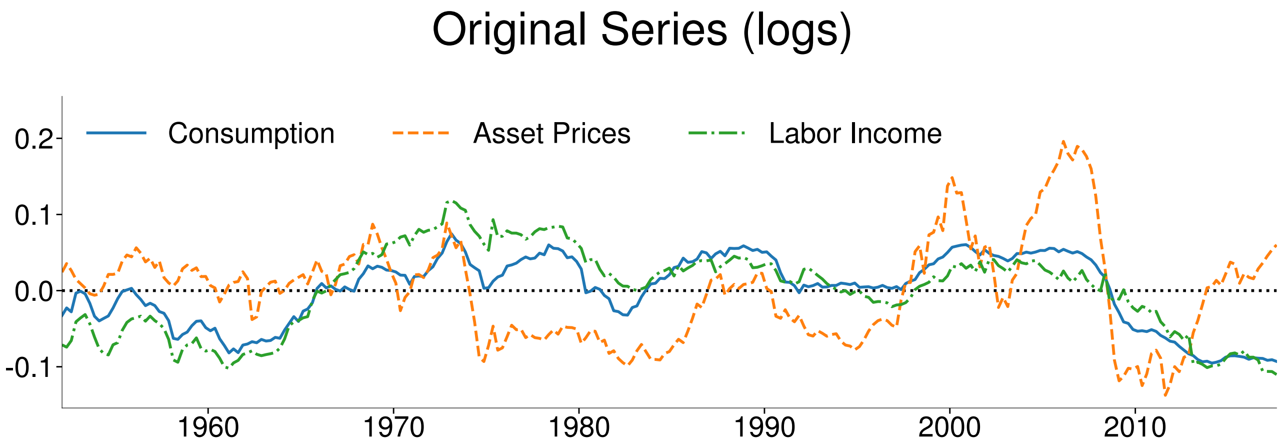
Example: cay

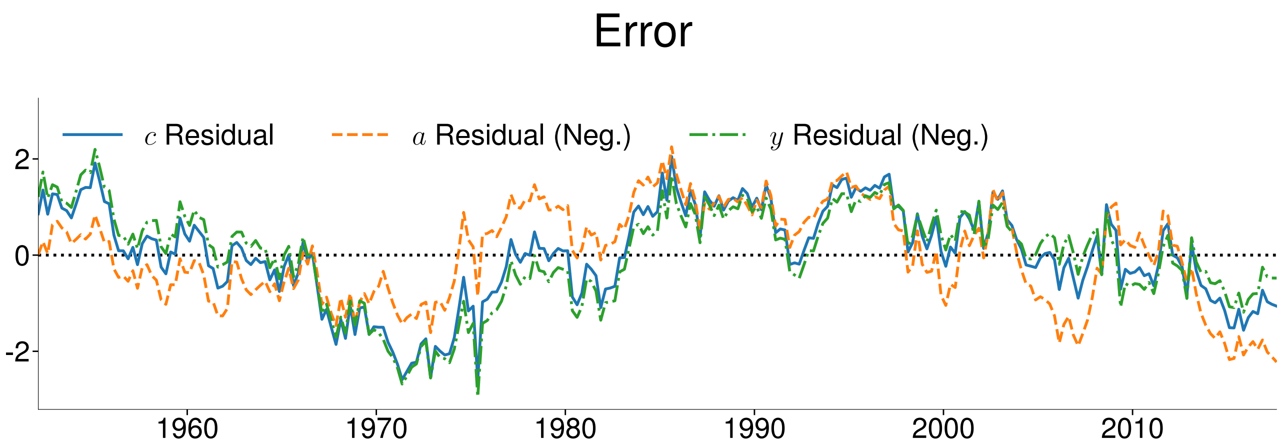
- Consumption-Aggregate Wealth has been an interesting cointegrated series in recent finance literature
- Has revived the CCAPM
- Three components:
  - Consumption (c)
  - Asset Wealth (a)
  - Labor Income (Human Wealth) (y)
- Deviation from long run related to expected return
- Cointegrating relationship: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

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

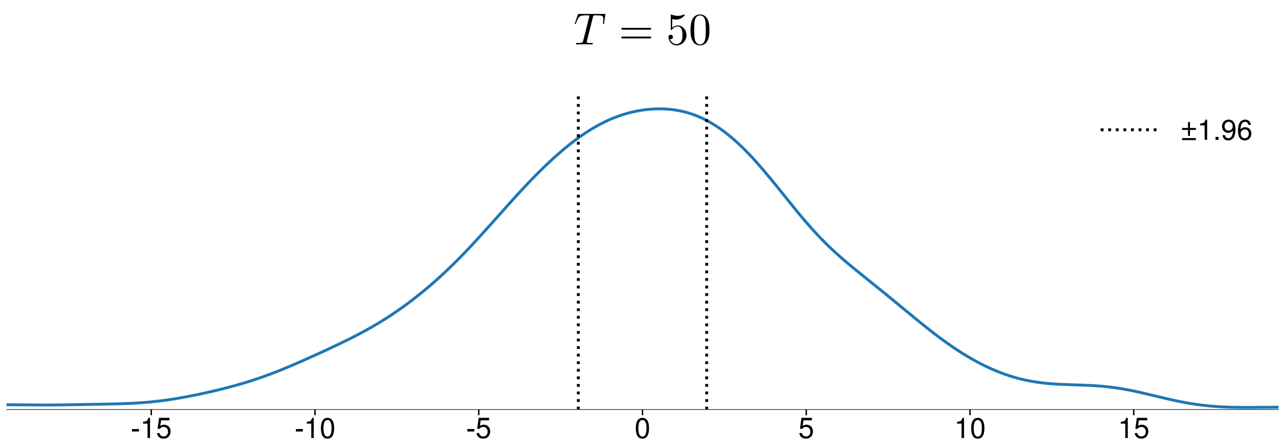
- Caution is needed when working with I(1) data
  - I(0) on I(0): The usual case. Standard asymptotic arguments apply.
  - I(1) on I(0): This regression is unbalanced.
  - I(1) on I(1): Cointegration or spurious regression.
  - I(0) on I(1): This regression is unbalanced.
- Spurious regression can lead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

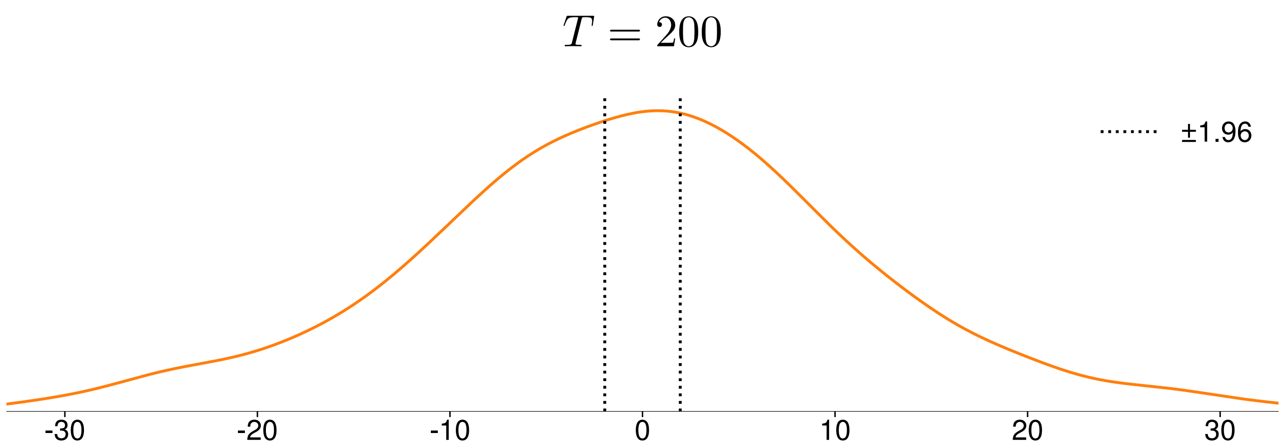
$$x_t = x_{t-1} + \epsilon_t$$

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

  - 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





- 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
  - $\mathrm{E}[\eta_t\eta_{t-1}] \neq 0$.

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

The case of an MA(1) error

- In terms of autocovariances,

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

- $\gamma_0 = \mathrm{V}[\eta_t]$=$\left(1+\theta^2\right)\mathrm{V}\left[\epsilon_t\right]$ 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,

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

and

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

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

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

where $\hat{\boldsymbol{\Gamma}}_l = T^{-1} \sum_{t=l+1}^{T} \mathbf{z}_t^* \mathbf{z}_{t-l}^{*\prime}$ 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